The Dissertation Committee for James Andrew Madsen Certifies that this is the approved version of the following dissertation:

Advancement of Photodissociation and Electron-Based Tandem Mass Spectrometry Methods for Proteome Analysis

Committee:

Jennifer S. Brodbelt, Supervisor

Kevin N. Dalby

Edward M. Marcotte

Lauren J. Webb

Katherine A. Willets

Advancement of Photodissociation and Electron-Based Tandem Mass Spectrometry Methods for Proteome Analysis

by

James Andrew Madsen, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2011

For Lindsey and Marsha

Acknowledgements

I am deeply indebted to family and friends for always supporting me in my varied, seemingly countless, endeavors. Thank you to both of my grandfathers, Dr. Samuel Gaston and Judge Harold Madsen, who showed me the importance of a formal education. Thank you to my brother, Michael, who is a solid example as to why success and intelligence are not necessarily determined by a fancy title. Thank you to my father, Bob, for being curious and truly interested in things (to some degree, these attributes have rubbed off on me). Thank you to my mother, Marsha, who got me interested in art at an early age, and fostered my creativity—I thank you for your unwavering support and selflessness.

I am grateful to my advisor, Jennifer Brodbelt, for your encouragement of innovative thinking, for your value of my input, for your tireless work ethic in turning around manuscript reviews in the shortest amount of time humanly possible, for your willingness to meet me halfway, for your open mind, and for your readiness to help—I progressed as an independent scientist far faster because of your support.

Many thanks to the members of the Brodbelt lab, past and present—your help in the lab has been invaluable. I have always enjoyed talking science and bouncing ideas off of all of you. The time we have spent outside of the lab has also been a real pleasure. It's been fun.

This dissertation would not have been possible without the collaboration among outside labs and organizations. I would like to thank Dr. Jae Schwartz and Dr. George Stafford, Jr. from ThermoFisher Scientific, and Dr. Joshua Coon and Aaron Ledvina from the University of Wisconsin at Madison for our successful collaboration in furthering the

capabilities of photodissociaiton. Thanks to Dr. Kevin Dalby and his graduate student, Tamer Kaoud, for providing interesting and applicable mitogen-activated pathway kinases. Thanks to Daniel Boutz for helping prepare and provide complicated cell samples and for our helpful discussions regarding proteomics. I would also like to thank (and apologize to) Dr. Edward Marcotte for borrowing Daniel, as well as some of his other students and resources, to utilize their expertise in various projects. I tried not to take up too much of their time, and I always offered to pay for the trypsin.

Finally, to my partner-in-crime / soul mate / personal designer / therapist, Lindsey Eden Turner, I thank you. This dissertation was so much easier because of you.

Advancement of Photodissociation and Electron-Based Tandem Mass Spectrometry Methods for Proteome Analysis

James Andrew Madsen, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Jennifer S. Brodbelt

The number and types of diagnostic ions obtained by infrared multiphoton dissociation (IRMPD) and collision induced dissociation (CID) were evaluated for supercharged peptide ions created by electrospray ionization of solutions spiked with *m*-nitrobenzyl alcohol. IRMPD of supercharged peptide ions increased the sequence coverage compared to that obtained by CID for all charge states investigated.

Multiply charged, N-terminally derivatized peptides were subjected to electron transfer reactions to produce singly charged, radical species. Upon subsequent "soft" CID, highly abundant *z*-type ions were formed nearly exclusively, which yielded simplified fragmentation patterns amenable to *de novo* sequencing methods. Furthermore, the simplified series of *z* ions were shown to retain labile phosphoric acid moieties.

Infrared multiphoton dissociation (IRMPD) was implemented in a novel dual pressure linear ion trap for rapid "top-down" proteomics. Due to secondary dissociation, IRMPD yielded product ions in significantly lower charge states as compared to CID, thus facilitating more accurate mass identification and streamlining product ion assignment. This outcome was especially useful for database searching of larger proteins (~29 kDa) as IRMPD substantially improved protein identification and scoring

confidence. Also, IRMPD showed an increased selectivity towards backbone cleavages N-terminal to proline and C-terminal to acidic residues (especially for the lowest precursor charge states).

Ultraviolet photodissociation (UVPD) at 193 nm was implemented on a linear ion trap mass spectrometer for high-throughput proteomic workflows. Upon irradiation by a single 5 ns laser pulse, efficient photodissociation of tryptic peptides was achieved with production of *a*, *b*, *c*, *x*, *y*, and *z* sequence ions, in addition to immonium ions and *v* and *w* side-chain loss ions. The factors that influence the UVPD mass spectra and subsequent *in silico* database searching via SEQUEST were evaluated.

193 nm ultraviolet photodissociation (UVPD) was employed to sequence singly and multiply charged peptide anions. Upon dissociation by this method, *a*-/*x*-type, followed by *d* and *w* side-chain loss ions, were the most prolific and abundant sequence ions, often yielding 100% sequence coverage. LC-MS/UVPD analysis using high pH mobile phases yielded efficient characterization of acidic peptides from mitogen-activated protein kinases.

Table of Contents

Chapter 1

Introduction

1.1 Introduction

Proteomics is the study of the structures and functional interplay of proteins in cells. By fully characterizing proteins and their extensive network of post-translational modifications (both qualitatively and quantitatively), scientists would have a better understanding and approach for developing new therapeutics that target cellular disease. Mass spectrometry (MS) has increasingly become the analytical method of choice in the field of proteomics in recent years due to significant advances in instrumentation, sampling techniques, and data interpretation algorithms. Although there have been tremendous inroads, there still remain many unresolved limitations in fully characterizing cellular proteomes. This dissertation aims at advancing the development of novel tandem mass spectrometry techniques (MS/MS and MS^n), a vital component of proteomic analysis, to increase the diagnostic content of MS/MS spectra, improve automated analysis, decrease experimental duty cycles, and overall furthering the field of proteomics by developing methods to maximize the characterization of proteins.

1.2 Tandem Mass Spectrometry for Proteomics

Large scale, "bottom-up" characterization of cellular proteomes has been enormously successful[1] and continues to improve due to ongoing innovation in the areas of mass spectrometric instrumentation, tandem mass spectrometry (MS^n) techniques, sampling, separations, and bioinformatics. Typical high-throughput, bottom-up workflows consist of proteins being extracted from cells, enzymatic digestion, separation

of the resulting peptides, and then online analysis by MS and MS/MS. Database search algorithms are then employed to correlate MS/MS spectra with sequence and ultimately the parent protein. Four of the most popular algorithms, SEQUEST,[2] MASCOT,[3] OMSSA,[4] and X! Tandem,[5] were recently compared,[6] and each algorithm yielded acceptable and similar results for a complex human protein sample. In terms of MS instrumentation, the introduction of the Orbitrap[7, 8] and hybrid linear ion trap (LIT)-Orbitraps[9, 10] have afforded significantly better mass accuracy and resolution (i.e., a resolving power of ~80,000 – 100,000 at *m/z* of around 1000)[7], which has greatly increased selectivity in database searching for bottom-up experiments. Recent innovations in quadrupole-time-of-flight instruments have yielded similar high performing mass accuracy and resolution measurements.[11, 12] Also, the analysis times of mass spectrometers continue to decrease. The newly released dual-pressure linear ion trap (Velos),[13] for example, has decreased cycle times two-fold by technological advances that eliminate prescans and allow faster scan rates in a low pressure trap. This improvement significantly increased the experimental duty cycle over more traditional trapping instruments, resulting in more protein identifications and a deeper depth of analysis into the proteome.[13]

While the most common proteomics technique for protein analysis uses the "bottom-up" mass spectrometry strategy,[14, 15] the "top-down" strategy (an alternative to the bottom-up approach) relies on fragmentation of intact protein ions in the gas phase (i.e., no enzymatic digestion). Although more challenging, top-down methods offer certain benefits over the bottom-up approaches including the elimination of enzymatic digestion procedures, access to contextual post-translational modification (PTM) information, and direct knowledge about the molecular weights of the intact proteins.[16-18] To date, top-down proteomic applications have mainly been undertaken on FT-ICR

instruments,[16] which can resolve highly charged product ions created upon collision induced dissociation (CID),[19-21] electron capture dissociation (ECD),[22-24] infrared multiphoton dissociation (IRMPD),[25] or ultraviolet photodissociation (UVPD)[26] of the intact proteins. Top-down applications utilizing various hybrid instrument platforms such as the linear ion trap (LIT)/FT-ICR,[17] LIT/time-of-flight,[18, 27] and LIT/orbitrap[28] that allow physical separation of the ion activation and mass analysis steps have gained momentum. Recently, some inroads in top-down proteomics have been achieved using cheaper, faster, and more robust quadrupole ion trap instruments.[12-18] For instance, McLuckey and coworkers have shown successful protein identification by dissociation of whole proteins by CID followed by gas-phase ion/ion reactions undertaken to convert multiply charged product ions into singly charged ions for easier product ion assignment.[29-32] Other groups have reported top-down protein analysis in quadrupole ion traps by combining electron transfer dissociation (ETD) with other ion/ion reactions[33] or by using extended ETD periods.[34] Although these ion trap techniques have shown substantial promise in the identification of proteins, they generally require relatively long activation/reaction times (up to 300 ms)[34] or require greater abundances of precursor ions than conventional CID.[33] Some of the resulting singly charged product ions that are useful for protein identification may also have *m/z* values beyond the range of a typical quadrupole ion trap (i.e., greater than *m/z* 2000) due to the subsequent proton transfer reactions.[33] Regardless if one uses the bottom-up or top-down strategy, the activation method of choice is critical to the success of the analysis, and more importantly, both strategies can be used in a complementary fashion to achieve the most complete characterization of the proteome.

1.2.1 Protein/Peptide Product Ion Nomenclature

To gain sequence information and to pinpoint modified amino acid residues in proteomics experiments, ions are fragmented into smaller diagnostic product ions. These fragmentation patterns can then be analyzed with automated algorithms to identify and characterize the peptide(s)/proteins(s) of interest. For any given activation method, the most important outcome succeeding dissociation is the generation of product ions that can be predicted. That is, for any given sequence, the peptide will fragment along the backbone and/or at the amino residues specifically and reproducibly. To simplify the interpretation of these fragmentation patterns, nomenclature was proposed and was accepted throughout the mass spectrometry and proteomics communities. This general nomenclature for peptide fragmentation along the backbone is illustrated in Figure 1.1.

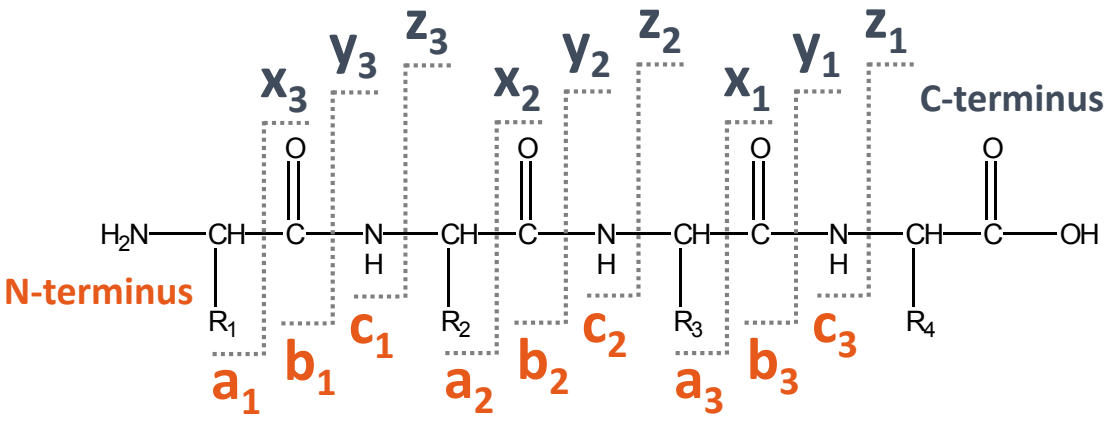

Figure 1.1 Peptide/Protein Fragmentation Nomenclature

The main products formed from all activation methods described in this dissertation are contained within this general nomenclature scheme. The ions are labeled by the specific cleavage site along the backbone and to which termini (N or C) the product contains.

Internal fragments, those that contain no termini, have no assigned nomenclature since they are generally produced with low frequency and abundance, which is true for almost all activation methods. Numbers are assigned to each product ion based off how many residues from the termini a particular cleavage occurs. Dissociation by thermal activation generally yield cleavages at N-C(O) bonds along the backbone producing *b*- and *y*-type products. In contrast, dissociation by higher-energy photons and collisions more often yield higher abundances of fragment ions from C_α-C(O) cleavages, which produce *a*/*x* ions. Lastly c- and z- type ions are generated from N- C_α backbone cleavage, and are generally the result of activation methods that utilize radical-directed mechanisms. Depending on the tandem MS method, losses of side-chain moieties from these product ions can occur at high frequencies and abundances. These losses occur from secondary dissociation mechanisms and specifically for high-energy activation methods; thus, a more detailed discussion for these ions will be discussed in section 1.2.2.4 of the Introduction (Vacuum Ultraviolet Photodissociaiton).

1.2.2 MS^n Methods

For successful protein identification, both bottom-up and top-down approaches rely on the collection of informative tandem mass spectra that critically depend on the activation technique used. In recent years there have been many advances in activation technology, but there still remains no universal method that can successfully and efficiently analyze all proteins and their peptide constituents. The activation methods that are utilized in the chapters herein will be discussed in the following sections.

1.2.2.1 Collision-Induced (Activated) Dissociation

Collision-induced dissociation (CID) is the gold standard MS/MS method. It essentially is the default activation method for almost all modern, commercial mass spectrometers. CID fragments precursor ions by accelerating them into a neutral gas (often Helium or Nitrogen) by applying an electrical potential. Some portion of the kinetic energy of the accelerated ions is then converted to internal energy thus causing dissociation from collisional heating. In most modern MS instruments used in proteomic workflows, these low-energy CID conditions (eV collisions) are employed in which dissociation occurs upon vibrational excitation generating *b*- and *y*-type product ions. Past mass spectrometers such as magnet sector instruments have employed high-energy (keV) CID conditions where the production of *a*-type, internal, and side-chain loss ions are frequent and abundant. However, this technique is not commonplace in modern proteomic analysis aside from TOF-TOF instruments.

Low-energy CID is an established, sensitive, and robust method, and when used to dissociate tryptic peptides, consistently identified the most peptides of complex protein mixtures as compared to other activation methods.[35, 36] Because of this, CID is the most common method used for sequencing peptides and identifying PTMs.[37] Uninformative labile neutral losses for both unmodified and modified peptides, however, remain a consistent problem for CID. Also, since internal energy is randomized prior to dissociation, collisional heating preferentially cleaves the backbone at specific residues (e.g., proline, histidine, glutamic acid, and aspartic acid) due to their lability, thus limiting the overall sequence coverage for certain peptides and essentially all top-down analysis. Another downfall of CID in ion traps, is the low mass cutoff (LMCO) problem, which arises due to an experimental incompatibility between the AC waveform used for CID and the radio frequency (rf) trapping voltage. This limitation prevents trapping and

detection of informative, low *m/z* product ions such as small *b* and *y* ions as well as immonium ions, the latter which are particularly useful in determining the amino acid composition of unknown peptides. Furthermore, these lower *m/z* diagnostic ions are important for automated sequencing algorithms and can be useful for the rapid determination of modified amino acids (e.g., phosphorylated, acetylated, and methylated residues).[38-40] Recently, however, higher-energy C-trap dissociation (HCD), a form of CID that allows smaller *m/z* product ions to be detected, has been implemented on the newest hybrid Orbitraps and has yielded comparable results to that of low-energy CID for complex biological samples.[10, 41] This new technique shows promise in both bottom-up and top-down proteomic applications, but its full exploitation has yet to be fully realized.

1.2.2.2 Infrared Multiphoton Dissociation

Infrared multiphoton dissociation (IRMPD) is another tandem MS method that has been successfully implemented in FT-ICR, time-of-flight, and quadrupole ion trap instruments.[25, 42-58] Much of the appeal of IRMPD in quadrupole ion traps is related to the broader *m/z* trapping range possible compared to conventional CID. This stems from the ability to reduce the rf trapping voltage during ion activation, an option that is detrimental for CID due to the decrease in energy deposition associated with lower rf trapping voltages.[57] Also, because all ions are continuously irradiated during the activation period, IRMPD can promote secondary dissociation of primary product ions and thus lead to formation of a more diverse array of diagnostic ions compared to CID. [57, 59] Due to the substantial competition between IR energization and collisional cooling that occurs at typical operating pressures of quadrupole ion traps (~1 mtorr), IRMPD has had limited implementation in quadrupole ion trap instruments. Several methods have been explored to combat the low IRMPD efficiencies of larger peptides. For example, the Glish group

developed thermally-assisted IRMPD (TA-IRMPD),[57] which involved heating the bath gas in the ion trap to increase the internal energies of ions, and have also improved dissociation efficiencies from simply focusing the laser.[60] Our group has pursued peptide derivatization strategies to increase the photoabsorptivity of ions and/or decrease their critical energies for dissociation, [58, 61-63] the supercharging of ions prior to activation,[64] and more recently the utilization of a dual pressure linear ion trap.[65]

Efficient top-down proteomics using IRMPD in quadrupole ion traps has remained largely unexplored. Due to the inherent lower pressures of FT-ICR and linear ion trap/time-of-flight mass analyzers, both McLafferty and Siu have used these instruments in conjunction with IRMPD for successful top-down analysis.[25, 66] In both of these prior studies, the dominant IRMPD pathways involved cleavages of the protein backbone on the C-terminal side of glutamic and aspartic acid residues as well as on the N-terminal side of proline residues.[25, 66]

1.2.2.3 Electron Transfer/Capture Dissociation

Electron-based dissociation techniques such as electron capture dissociation (ECD)[22] in FT-ICR instruments and electron transfer dissociation (ETD)[67] in ion traps have shown to be particularly promising for characterization of post-translational modifications (PTMs)[44, 67-73] and for increasing peptide sequence coverage.[35, 74] Both methods involve the reaction of multiply charged cations with either low-energy electrons (ECD) or radical anions (ETD), leading to an exothermic electron attachment process that may also cause complementary *c*- and *z*-type backbone cleavages that retain labile PTMs such as phosphorylation and glycosylation. In more mechanistic detail, electron capture/transfer dissociation requires the migration of an H atom to a carbonyl group in the peptide backbone prior to cleavage of an N-C_α bond and production of

complementary *c* and $z^{\bullet}$ ions.[22, 67] This process is thought to involve rearrangement and transfer of α-carbon radicals to backbone carbonyls, initiating a free radical reaction cascade.[75] To a lesser extent an H atom can be transferred to an amide nitrogen, thus generating *a*/*y* fragments.[67] For peptides of lower charge, however, these electron-based methods yield charge-reduced radicals from the precursor as the most abundant products with low yields of *c* and *z* products.[74, 76] Recently, both the Coon and McLuckey groups have implemented supplemental collisional activation (i.e. low-energy CID) of charge-reduced peptide radicals after the electron transfer reaction (ETcaD), thus affording significantly improved peptide sequence coverage for lower charged precursor ions.[76-78] Both ETD and ECD are promising new alternatives to CID for sequencing peptides.

1.2.2.4 Vacuum Ultraviolet Photodissociation

Vacuum ultraviolet photodissociation (UVPD) is another emerging activation method suitable for energizing and dissociating peptides. UVPD, which typically uses an excimer laser operated at a wavelength of 193 or 157 nm, delivers high-energy photons (6.4 or 7.9 eV) that can excite peptides to higher electronic states and/or promote access to other types of fragmentation pathways such as C_{α}-C(O) bond cleavages, resulting in *a* + 1 or *x* + 1 radical ions that may further undergo secondary dissociation into even electron *a*, *x*, *d*, *w*, *v*, and/or *Y* products through losses of CO and various amino acid side-chains.[79, 80] This secondary dissociation process is illustrated in **Figure 1.2**. The interesting fragmentation behavior arising from VUVPD has been exploited recently for proteomic applications.[80]

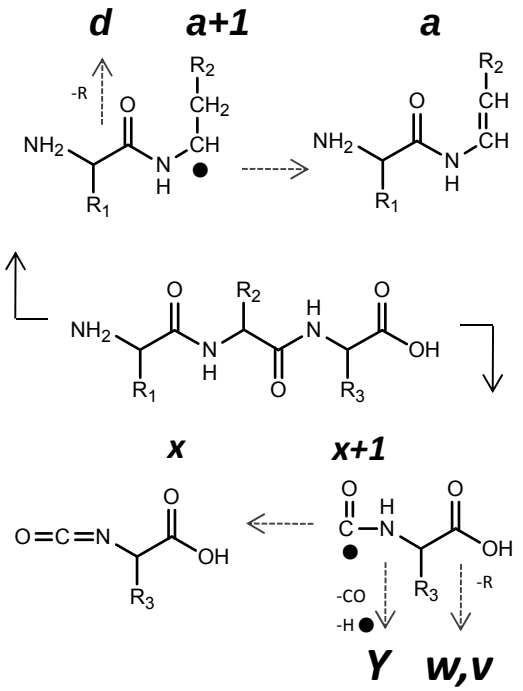

Figure 1.2 The main ions created upon UVPD of peptide cations. Dashed arrows indicate indirect pathways to other secondary product ions.

UVPD at 193 nm of peptides was first performed in the 1980's,[81-84] but experiments were limited to a few selected peptides on FTICR mass spectrometers. A recent renaissance in the use of short wavelength (157 and 193 nm)[85-90] ultraviolet photodissociation (UVPD) coupled mainly to TOF instruments has occurred due to the ultrafast (nanosecond) activation timescales and the rich tandem mass spectrometric information obtained (i.e., more types of backbone and side-chain loss ions are observed compared to traditional activation methods). There have been noteworthy inroads in the implementation of MALDI-TOF-MS with UVPD for high-throughput proteomic workflows,[32] although inefficient ion activation has hindered the analysis of the lowest abundance proteins/peptides,[86] coupled with the typical need for ~2000 spectral averages

(~40 s) for each spectrum.[80, 85] The rich dissociation observed upon UVPD motivated our interests in implementing this activation technique in trapping instruments for high-throughput proteomic experiments.

1.3 Overview of Chapters

With all the innovations in tandem mass spectrometry that have helped advance the field of proteomics, there are still vast improvements that can be made to further the depth of knowledge about cellular proteomes. The body of work in this dissertation has sought to develop new methods and/or improve existing techniques to expand MS^n technology to its furthest potential.

In chapter 3, diagnostic ions obtained by IRMPD and CID were evaluated for supercharged peptide ions created by electrospray ionization of solutions spiked with *m*-nitrobenzyl alcohol. IRMPD of supercharged peptide ions increased the sequence coverage compared to that obtained by CID for all charge states investigated.

Chapter 4 utilized electron transfer reactions to produce singly charged, radical species from multiply charged, N-terminally derivatized peptides. Upon subsequent collision activation (MS^3), fragmentation patterns amenable to *de novo* sequencing were produced from a simplified series of *z*-type ions.

IRMPD was implemented in a novel dual pressure linear ion trap for rapid "top-down" proteomics in chapter 5. Due to secondary dissociation, product ions in significantly lower charge states were produced, thus facilitating more accurate mass identification and streamlining product ion assignment. This outcome was especially useful for database searching as IRMPD substantially improved protein identification and scoring confidence.

In chapter 6, UVPD at 193 nm was employed for high-throughput proteomic workflows. Upon irradiation by a single 5 ns laser pulse, efficient photodissociation was achieved yielding diagnostic fragmentation behavior from several complementary product ion types. The factors that influence the UVPD mass spectra and subsequent *in silico* database searching via SEQUEST were evaluated.

The results in chapter 7 highlighted the utility of 193 nm UVPD for sequencing singly and multiply charged peptide anions. Upon dissociation by this method, *a*-/*x*-type, were the most prolific and abundant sequence ions, often yielding 100% sequence coverage. LC-MS/UVPD analysis using high pH mobile phases yielded efficient characterization of acidic peptides from mitogen-activated protein kinases.

1.3 References

(1) de Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Froehlich, F.; Walther, T. C.; Mann, M. *Nature* **2008**, *455*, 1251-1254.

(2) Eng, J. K.; McCormack, A. L.; Yates, J. R., III *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 976-989.

(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

(4) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *Journal of Proteome Research* **2004**, *3*, 958-964.

(5) Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466-1467.

(6) Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. *Molecular and Cellular Proteomics* **2007**, *6*, 1599-1608.

(7) Hardman, M.; Makarov, A. A. *Analytical Chemistry* **2003**, *75*, 1699-1705.

(8) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. *Journal of Mass Spectrometry* **2005**, *40*, 430-443.

(9) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Analytical Chemistry* **2006**, *78*, 2113-2120.

(10) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. *Molecular and Cellular Proteomics* **2009**, *8*, 2759-2769.

(11) Hartmer, R. G.; Kaplan, D. A.; Stoermer, C.; Lubeck, M.; Park, M. A. *Rapid Communications in Mass Spectrometry* **2009**, *23*, 2273-2282.

(12) Ow, S. Y.; Noirel, J.; Salim, M.; Evans, C.; Watson, R.; Wright, P. C. *Proteomics* **2010**, *10*, 2205-2213.

(13) Pekar Second, T.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. *Analytical Chemistry* **2009**, *81*, 7757-7765.

(14) Chait, B. T. *Science* **2006**, *314*, 65-66.

(15) Collier, T. S.; Hawkridge, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C. *Analytical Chemistry* **2008**, *80*, 4994-5001.

(16) Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.

(17) Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne, M. T., II; Burke, P. V.; Kwast, K. E.; Kelleher, N. L. *Analytical Chemistry* **2007**, *79*, 7984-7991.

(18) Liu, J.; Huang, T.-Y.; McLuckey, S. A. *Analytical Chemistry* **2009**, *81*, 1433-1441.

(19) Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M. W.; McLafferty, F. W. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 286-289.

(20) Senko, M. W.; Speir, J. P.; McLafferty, F. W. *Analytical Chemistry* **1994**, *66*, 2801-2808.
(21) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *Journal of the American Chemical Society* **1999**, *121*, 806-812.
(22) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.
(23) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Analytical Chemistry* **2000**, *72*, 563-573.
(24) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 10313-10317.
(25) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Analytical Chemistry* **1994**, *66*, 2809-2815.
(26) Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little Daniel, P.; McLafferty, F. W. *International Journal of Mass Spectrometry and Ion Processes* **1996**, *157/158*, 357-364.
(27) Liu, J.; Huang, T.-Y.; McLuckey, S. A. *Analytical Chemistry* **2009**, *81*, 2159-2167.
(28) Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. *Molecular and Cellular Proteomics* **2006**, *5*, 949-958.
(29) Amunugama, R.; Hogan, J. M.; Newton, K. A.; McLuckey, S. A. *Analytical Chemistry* **2004**, *76*, 720-727.
(30) Hogan, J. M.; Pitteri, S. J.; McLuckey, S. A. *Analytical Chemistry* **2003**, *75*, 6509-6516.
(31) Reid, G. E.; McLuckey, S. A. *Journal of Mass Spectrometry* **2002**, *37*, 663-675.
(32) Bowers, J. J.; Liu, J.; Gunawardena, H. P.; McLuckey, S. A. *Journal of Mass Spectrometry* **2008**, *43*, 23-34.
(33) Coon, J. J.; Ueberheide, B.; Syka, J. E. P.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 9463-9468.
(34) Bunger, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L., Jr. *Analytical Chemistry* **2008**, *80*, 1459-1467.
(35) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. *Analytical Chemistry* **2008**, *80*, 4825-4835.
(36) Swaney, D. L.; Wenger, C. D.; Coon, J. J. *Journal of Proteome Research* **2010**, *9*, 1323-1329.
(37) Wells, J. M.; McLuckey, S. A. *Methods in Enzymology* **2005**, *402*, 148-185.
(38) Steen, H.; Fernandez, M.; Ghaffari, S.; Pandey, A.; Mann, M. *Molecular & cellular proteomics : MCP* **2003**, *2*, 138-145.
(39) Trelle, M. B.; Jensen, O. N. *Analytical Chemistry* **2008**, *80*, 3422-3430.
(40) Couttas, T. A.; Raftery, M. J.; Bernardini, G.; Wilkins, M. R. *Journal of Proteome Research* **2008**, *7*, 2632-2641.

(41) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nature Methods* **2007**, *4*, 709-712.
(42) Li, W.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. *Analytical Chemistry* **1999**, *71*, 4397-4402.
(43) Hofstadler, S. A.; Sannes-Lowery, K. A.; Griffey, R. H. *Analytical Chemistry* **1999**, *71*, 2067-2070.
(44) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Analytical Chemistry* **2001**, *73*, 4530-4536.
(45) Flora, J. W.; Muddiman, D. C. *Analytical Chemistry* **2001**, *73*, 3305-3311.
(46) Flora, J. W.; Muddiman, D. C. *Journal of the American Chemical Society* **2002**, *124*, 6546-6547.
(47) Drader, J. J.; Hannis, J. C.; Hofstadler, S. A. *Analytical Chemistry* **2003**, *75*, 3669-3674.
(48) Hofstadler, S. A.; Drader, J. J.; Gaus, H.; Hannis, J. C.; Sannes-Lowery, K. A. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 1413-1423.
(49) Chalmers, M. J.; Quinn, J. P.; Blakney, G. T.; Emmett, M. R.; Mischak, H.; Gaskell, S. J.; Marshall, A. G. *Journal of Proteome Research* **2003**, *2*, 373-382.
(50) Flora, J. W.; Muddiman, D. C. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 121-127.
(51) Chalmers, M. J.; Hakansson, K.; Johnson, R.; Smith, R.; Shen, J.; Emmett, M. R.; Marshall, A. G. *Proteomics* **2004**, *4*, 970-981.
(52) Tsybin, Y. O.; Ramstroem, M.; Witt, M.; Baykut, G.; Hakansson, P. *Journal of Mass Spectrometry* **2004**, *39*, 719-729.
(53) Yamada, N.; Suzuki, E.-I.; Hirayama, K. *Analytical Biochemistry* **2006**, *348*, 139-147.
(54) Mihalca, R.; van der Burgt, Y. E. M.; McDonnell, L. A.; Duursma, M.; Cerjak, I.; Heck, A. J. R.; Heeren, R. M. A. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 1838-1844.
(55) Clowers, B. H.; Dodds, E. D.; Seipert, R. R.; Lebrilla, C. B. *Journal of Proteome Research* **2007**, *6*, 4032-4040.
(56) Colorado, A.; Shen, J. X.; Brodbelt, J. *Analytical Chemistry* **1996**, *68*, 4033-4043.
(57) Payne, A. H.; Glish, G. L. *Analytical Chemistry* **2001**, *73*, 3542-3548.
(58) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80*, 4807-4819.
(59) Crowe, M. C.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 1581-1592.
(60) Newsome, G. A.; Glish, G. L. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1127-1131.
(61) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 6855-6862.
(62) Pikulski, M.; Wilson, J. J.; Aguilar, A.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 8512-8517.
(63) Pikulski, M.; Hargrove, A.; Shabbir, S. H.; Anslyn, E. V.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 2094-2106.

(64) Madsen, J. A.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 349-358.
(65) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C., Jr.; Brodbelt, J. S. *Analytical Chemistry* **2009**, submitted.
(66) Raspopov, S. A.; El-Faramawy, A.; Thomson, B. A.; Siu, K. W. M. *Analytical Chemistry* **2006**, *78*, 4572-4577.
(67) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(68) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Analytical Chemistry* **1999**, *71*, 4431-4436.
(69) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 1793-1800.
(70) Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 7450-7456.
(71) Srikanth, R.; Wilson, J.; Bridgewater, J. D.; Numbers, J. R.; Lim, J.; Olbris, M. R.; Kettani, A.; Vachet, R. W. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1499-1506.
(72) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. *Molecular and Cellular Proteomics* **2008**, *7*, 1838-1849.
(73) Sweet, S. M. M.; Mardakheh, F. K.; Ryan, K. J. P.; Langton, A. J.; Heath, J. K.; Cooper, H. J. *Analytical Chemistry* **2008**, *80*, 6650-6657.
(74) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 9243-9252.
(75) Leymarie, N.; Costello, C. E.; O'Connor, P. B. *Journal of the American Chemical Society* **2003**, *125*, 8949-8958.
(76) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Analytical Chemistry* **2007**, *79*, 477-485.
(77) Han, H.; Xia, Y.; McLuckey, S. A. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 1567-1573.
(78) Xia, Y.; Han, H.; McLuckey, S. A. *Analytical Chemistry* **2008**, *80*, 1111-1117.
(79) Zhang, L.; Reilly, J. P. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1378-1390.
(80) Zhang, L.; Reilly, J. P. *Journal of Proteome Research* **2010**, *9*, 3025-3034.
(81) Bowers, W. D.; Delbert, S. S.; Hunter, R. L.; McIver, R. T., Jr. *Journal of the American Chemical Society* **1984**, *106*, 7288-7289.
(82) Hunt, D. F.; Shabanowitz, J.; Yates, J. R., III *Journal of the Chemical Society, Chemical Communications* **1987**, 548-550.
(83) Hunt, D. F.; Shabanowitz, J.; Yates, J. R., III; Griffin, P. R.; Zhu, N. Z. *Analytica Chimica Acta* **1989**, *225*, 1-10.
(84) Lebrilla, C. B.; Wang, D. T. S.; Mizoguchi, T. J.; McIver, R. T., Jr. *Journal of the American Chemical Society* **1989**, *111*, 8593-8598.

(85) Zhang, L.; Reilly, J. P. *Analytical Chemistry* **2010**, *82*, 898-908.
(86) Kim, T.-Y.; Reilly, J. P. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 2334-2341.
(87) Shin, Y. S.; Moon, J. H.; Kim, M. S. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 53-59.
(88) Thompson, M. S.; Cui, W.; Reilly, J. P. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1439-1452.
(89) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 1657-1665.
(90) Morgan, J. W.; Russell, D. H. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 721-729.

Chapter 2

Experimental Methods

2.1 Mass Spectrometry

All mass spectrometric analysis in the chapters herein utilized electrospray ionization (ESI) and ion trap mass spectrometers. ESI is the most commonly used ionization technique in proteomics research due to its ability to gently ionize peptides/proteins without extensive degradation and denaturing. To achieve ionization, a high voltage of approximately 2 – 5 kilovolts is applied to a sample solution inside a narrow capillary; upon sufficient flow, a charged spray is formed containing the ionized analyte of interest. ESI generates ions that are multiply charged, which is advantageous since the most successful ion activation methods yield the most informative fragmentation information from multiply charged precursor ions. The majority of the work herein focuses on the development of novel ion activation methods that either provide more extensive and diagnostic fragmentation patterns for proteomic analysis, lowers the activation times and improves the duty cycle of LC-MS/MS experiments, and/or yield more easily interpreted MS/MS patterns that may be useful for automated spectral interpretation algorithms. The mass spectrometric instrumentation utilized in this body of work is described in detail in the subsequent sections.

2.1.1 Thermo Electron LTQ Ion Trap Mass Spectrometer

Peptides were analyzed using a modified Thermo Fisher (San Jose, CA) LTQ two-dimensional linear ion trap mass spectrometer. Pulsed-Q Dissociation (P-QD), a

form of CID, was performed on the LTQ with a q value of 0.7 and an activation time of 0.1 ms.

Direct infusion nanoESI experiments utilized an online setup using a conductive mini microfilter assembly from IDEX Health and Science (Oak Harbor, WA) coupled to a New Objectives uncoated PicoTip® nanoESI emitter (Woburn, MA), which was operated at an ESI voltage of 2 kV. Peptides diluted to 10 μM were directly infused by a syringe pump at 300 nL/min.

2.1.2 ThermoFinnigan LCQ Ion Trap Mass Spectrometer

Supercharged tryptic peptides were analyzed using a modified ThermoFinnigan LCQ Deca XP three-dimensional quadrupole ion trap mass spectrometer equipped with a Synrad 50-W continuous wave CO_2 laser.[1] Radiation at 10.6 μm was introduced into the quadrupole ion trap through a 5 mm hole drilled into the ring electrode. For certain tryptic peptides, the He bath gas pressure was reduced during laser dissociation experiments to allow for greater fragmentation efficiencies. A higher q value of 0.25 was used for CID experiments to afford more efficient ion activation at the expense of a more limited *m/z* trapping range. For all charge states, the CID voltage or IR irradiation time was adjusted to cause dissociation of approximately 100% (if possible) of the precursor ions.

2.1.3 Thermo Electron LTQ Velos Ion Trap Mass Spectrometer

Proteins were analyzed using a modified LTQ Velos, which is a differentially pumped dual pressure linear ion trap (Thermo Fisher, San Jose, CA).[2] The high pressure cell was operated at nominally 4.7 mTorr, and the low pressure cell at nominally 0.3 mTorr. Protein ions are accumulated in the high pressure cell for more efficient trapping

and isolation, and were then transferred to the low pressure cell where m/z analysis was performed in zoom scan mode. The automatic gain control (AGC) was set to 500, and the ESI voltage to 4.5 kV. Sheath and auxiliary nitrogen gas flow rates were 0.12 L/min and 0.3 L/min, respectively. The heated capillary temperature was maintained at 200ºC. IRMPD experiments were performed in both the high and low pressure cells using a Synrad 50 W continuous wave CO_2 laser (model 48-5; Mukilteo, WA) operated at full power (~50 W). Laser radiation at a wavelength of 10.6 μm was introduced into the linear ion trap via a modified back plate containing a ZnSe window.[3] The precursor m/z was placed at a q-value of 0.1 for all IRMPD experiments, and irradiation times ranged from 5.0 – 18.0 ms. CID experiments were performed in the high pressure cell to maximize dissociation efficiencies. Standard parameters (q-value of 0.25 and activation times of 30 ms) were used for all CID experiments. All spectra were generated from 20 – 34 averages of three microscans.

2.2 Liquid Chromatography

For complex proteomic samples, often a separation step is required prior to mass spectrometric analysis. This step reduces ion suppression because the complex mixture of analytes is dispersed in time which prevents saturation of the electrospray, and increases the richness of the proteomic data. Most commonly, peptide/protein mixtures are separated via reversed phase - high performance liquid chromatography (RP-HPLC), which is then easily coupled online to ESI-MS/MS instrumentation. The HPLC flow rate is a critical parameter to the success of LC-MS/MS and can be divided into three commonly used flow rate scales - analytical (mL flow rates), capillary (μL flow rates), and nano (nL flow rates). The lower the flow rate, the more sensitive the LC-MS/MS method is in terms of total ion intensity; however, these low flow rate regimes are often

not robust and require specialized instrumentation. The HPLC instrumentation and parameters that were utilized in the chapters herein are described in the subsequent sections.

2.2.1 Hitachi L-7000 HPLC

Samples containing 10 μM digested protein were injected (10 μL) onto a Symmetry300 reversed-phase C_{18} analytical column (Waters, Milford, MA) (2.1 × 50 mm, 3.5 μm packing) with a matched guard column (2.1 × 10 mm, 3.5 μm packing). Eluents consisted of 0.2% formic acid in water (A), and 0.2% formic acid in acetonitrile (B). Gradient elution was performed as follows: 95% (A) for 2 minutes followed by a linear gradient to 40% (A) over 60 minutes at a flow rate of 0.300 mL/min. Data-dependent aquisition was used for all LC-MS^n analysis with the first scan event being a full mass spectrum at a *m/z* range of 400 – 2000, and subsequent events consisting of ETD, CID, and ETcaD of the one or two most abundant peaks. A normalized collision energy of 35% was used for all LC-CID-MS analysis.

2.2.2 Dionex Ultimate 3000 cap/nano HPLC

Liquid chromatography was performed using a Dionex UltiMate 3000 system (Sunnyvale, CA) using a capillary flow splitter. An Agilent ZORBAX 300SB-C_{18} capillary column (Santa Clara, CA) (150 × 0.3 mm, 5 μm particle size) was used for all separations. The column temperature was kept constant at 40°C. Eluent A consisted of 0.1% formic acid in water and eluent B 0.1% formic acid in acetonitrile. A linear gradient from 5% eluent B to 40% eluent B over 65 min at 5 μL/min was used. Injections of approximately 1 μg (20 picomoles) were used for the digested BSA sample and 10 μg for the HT-1080 lysate sample.

Liquid chromatography using high pH eluents was carried out using a Dionex UltiMate 3000 system (Sunnyvale, CA) with an Agilent ZORBAX 300Extend-C_{18} capillary column (Santa Clara, CA) (150 × 0.3 mm, 3.5 μm particle size). Eluent A was composed of 10 mM piperidine (~pH 11.5) in water and eluent B 10 mM piperidine (~pH 11.5) in acetonitrile. A 65 min linear gradient from 5% eluent B to 40% eluent B was used at a flow rate of 4 μL/min. The mitogen-activated kinase mixture was injected at approximately five picomoles of tryptic digest per protein.

2.3 Chemicals

Myoglobin and cytochrome c from equine (*Equus caballus*) heart, bovine (*Bos Taurus*) carbonic anhydrase, bovine serum albumin, bovine ubiquitin, bovine α-SI-casein, proteomics grade trypsin, and all other reagents and peptides were purchased from Sigma-Aldrich (St. Louis, MO). The peptides DRVYIHPFHLVIH, YRPPGFSPFR, KRPPGFSPFR, ASHLGLAR, DRVYIHPFHLVIH, TSTEPQpYQPGENL, HCKFWW, GNHWAVGHLM, and DAEFRHDSGYEVHHEK were purchased from BACHEM (King of Prussia, PA). The phosphorylated peptides RQpSVELHSPQSLPR, KRpTIRR, TRDIYETDYpYRK and NRVpYIHPF were purchased from from AnaSpec (San Jose, Ca), and immobilized TPCK trypsin beads were obtained from Pierce Biotechnology, Inc. (Rockford, IL). KAKAA and RAAAA were synthesized by Bio Basic (Ontario, Canada). All solvents were purchased from Fisher Scientific (Fairlawn, NJ).

2.4 Protein Preparation

For TPCK trypsin bead digestion, proteins were digested with 100 μL immobilized TPCK trypsin beads in 50 mM ammonium bicarbonate incubated for 18

hours at 37 °C. After digestion, the excess beads were discarded and the remaining tryptic peptides were diluted prior to analysis.

Proteins with disulfide linkages were reduced by adding a 10-fold molar excess of dithiothreitol (DTT) (100 µL of 1.0 mM DTT) to 10 nmoles of BSA (10 µL of 1.0 mM protein), which was buffered to a pH of ~8.4 with ~4 µmoles ammonium bicarbonate. Solutions were then incubated for one hour at 40 ˚C after which alkylation was performed by adding 4.0 µL of 1.0 M iodoacetamide (buffered to a pH of ~8.0 with ~20 µmoles ammonium bicarbonate). This solution was incubated at ambient temperature for 45 minutes in the dark followed by subsequent quenching by the addition of excess DTT. For enzymatic digestion, 5.0 µL of 1.0 mg/mL trypsin in 1.0 mM HCl was added to the reduced/alkylated solution followed by incubation at 37˚C for 16 hours.

HT1080 cells were suspended in low-salt buffer (10 mM Tris-HCl, 10 mM KCl, 1.5 mM $MgCl_2$, pH 8.0) to swell, and were then lysed by dounce homogenization. The whole cell lysate was centrifuged at 1000 × g to clarify the soluble lysate and to remove the insoluble pellet. Soluble lysate (2 mg/mL protein) was denatured by addition of 2,2,2-trifluoroethanol (TFE) to 50% (vol/vol) and reduced with 15 mM DTT at 55 ˚C for 45 min. Following reduction, the proteins were alkylated with 55 mM iodoacetamide at room temperature for 30 min. The sample was diluted in digestion buffer (50 mM Tris-HCl, 2 mM $CaCl_2$, pH 8.0) to a final TFE concentration of 5% (vol/vol). Trypsin was added to a 1:50 (enzyme:protein) concentration (w/w), and the sample was incubated at 37˚C for 5 hrs. The digestion was quenched with 1% formic acid and sample volume was reduced to 20 µl by centrifugation under reduced pressure using a SpeedVac. Digested peptides were bound and washed on Thermo Fisher Scientific HyperSep C-18 SpinTips (San Jose, CA), resuspended in peptide buffer (95% H_2O/5% acetonitrile/0.1%

formic acid) and filtered through Millipore Microcon 10 kDa centrifugal filters (Billerica, MA), with the digested peptides collected as flow-through.

2.5 Peptide Derivatization

N-terminal derivatization was utilized in this dissertation to produce simplified MS/MS patterns for aiding *de novo* sequencing. Prior to N-terminal derivatization, lysines were converted to homoarginine residues through a guanidination step (except where noted) to ensure modification with 4-sulfophenyl isothiocyanate (SPTIC) exclusively at the N-terminus and not at the ε-amine group of lysines. A stock of 0.05 g O-methylisourea in 50 µL water was prepared and 5 µL of this stock solution, 50 µL of 5 N ammonium hydroxide, and approximately 10 nmol of the protein digest or model peptide was combined. This mixture was then incubated for 10 minutes at 65 °C, after which the sample was desalted via C_{18} spin columns.

The SPITC and 4-(chlorosulfonyl)phenyl isocyanate (SPC) derivatization procedures used in this study were adapted from previous reports.[4, 5] Briefly, a stock solution of SPITC was made by dissolving approximately 1 mg of SPITC in 100 µL of a 20 mM $NaHCO_3$ solution (pH ~9.5). A 20 µL aliquot of this stock was combined with 10 nmol of peptide or protein digest, which was then incubated for approximately 30 minutes at 55 °C. For SPC-modified peptides, a stock solution was made by dissolving approximately 1 mg of SPC in 100 µL acetonitrile. A 20 µL aliquot of this stock was combined with 10 nmol of peptide in a 50:50 pyridine/water (v/v) solution, which was then incubated for approximately 30 minutes at ambient temperature. After incubation, all samples were cleaned up using C_{18} spin columns. For direct infusion experiments, working solutions of 10 µM peptides in 49.5:49.5:1.0 water/methanol/acetic acid (v/v/v)

solution were prepared before infusion at a flow rate of 3 μL/min into the mass spectrometer.

2.6 Infrared Multiphoton Dissociation

Infrared multiphoton dissociation (IRMPD) was performed on a modified Thermo Fisher (San Jose, CA) LTQ two-dimensional linear ion trap mass spectrometer equipped with a Synrad 50-W continuous wave CO_2 laser (model 48-5; Mukilteo, WA). Laser radiation at a wavelength of 10.6 μm was introduced into the linear ion trap via a modified back plate containing a ZnSe window.[3] IRMPD was also utilized on a modified ThermoFinnigan LCQ Deca XP three-dimensional quadrupole ion trap mass spectrometer equipped with the same Synrad 50-W continuous wave CO_2 laser.[1] Radiation at 10.6 μm was introduced into the quadrupole ion trap through a 5 mm hole drilled into the ring electrode. With the laser operated at full power, irradiation times of 10 – 250 ms (LTQ) and 25 – 500 ms (LCQ) were used with an activation q value of 0.1. IRMPD in a modified Thermo Fisher (San Jose, CA) LTQ Velos was performed in both the high and low pressure cells using a Synrad 50 W continuous wave CO_2 laser (model 48-5; Mukilteo, WA) operated at full power (~50 W). The precursor m/z was placed at a q-value of 0.1 for all IRMPD experiments, and irradiation times ranged from 5.0 – 18.0 ms.

2.7 Electron Transfer Dissociation

Electron transfer dissociation (ETD) was performed on a Thermo Fisher (San Jose, CA) LTQ XL two-dimensional linear ion trap mass spectrometer equipped with an ETD unit. Fluoranthene anions were introduced as the electron-transfer reagent for ETD experiments with reaction times of 100 ms. ETcaD was carried out using ET reaction

times of 100 ms followed by isolation of the charge-reduced radical, and a 30 ms CID period.

2.8 Vacuum Ultraviolet Dissociation

Vacuum Ultraviolet Dissociation (UVPD) was performed on a Thermo Fisher Scientific LTQ XL linear ion trap (LIT) mass spectrometer (San Jose, CA) outfitted with a Coherent ExciStar XS excimer laser (Santa Clara, CA) operated at 193 nm. The laser setup was similar to that previously described,[1] except a CaF_2 lens was used to transmit 193 nm photons into the LIT. For UVPD experiments, the laser was pulsed once per scan with an energy of 8 mJ/pulse and a pulse duration of 5 ns. Pulse variable experiments used a laser pulse frequency of 500 Hz. Data-dependent LC-MS/MS was performed in two different ways. For LC-MS/UVPD runs (i.e., only UVPD was used for activation), the first event was the full mass scan (*m/z* range of 400 – 2000) followed by ten consecutive UVPD events on the ten most abundant ions from the full mass scan with one UV pulse applied per MS/MS scan. A *q*-value of 0.1 was used for each UVPD event. For LC-MS/UVPD/CID runs (i.e., comparison between UVPD and CID), the first event was the full mass scan (*m/z* range of 400 – 2000) followed by ten alternating UVPD/CID events on the five most abundant peaks for a total of five UVPD and five CID events per cycle. A *q*-value of 0.1 and a default activation period of 30 μs were used for each UVPD event. Although the actual irradiation time was only 5 ns (the duration of a single laser pulse) for UVPD, the commercial LTQ software limited the activation period to a minimum of 30 μs.

For activated photoelectron detachment dissociation (a-EPD) experiments, a *q*-value of 0.1 and an activation of 9 pulses (18 ms), 1 mJ/pulse, 500 Hz were used to generate the charge reduced, radical anion species, which was subsequently activated

using a normalized collision energy of 35%, *q*-value of 0.25, and collision duration of 30 ms.

2.9 Automated Database Searching

The automated searching of LC-MS/MS data is vital to the success of proteomic analysis. Typically, experimental MS/MS spectra are searched against theoretical spectra generated *in silico* from known protein sequence databases. Various scoring algorithms are then applied to the data to access its validity. False discovery rates (FDR) calculated from searching data sets against a reversed decoy database are also applied to further filter the true positives from the false positives. The database searching algorithms (both for intact and digested proteins) that were utilized in the chapters herein are described in the subsequent sections.

2.9.1 ProsightPTM 2.0

ProsightPTM 2.0[6-8] was used for database searching of *Bos taurus* proteins (e.g., ubiquitin and carbonic anhydrase). Peaks were picked by Origin 7.0, and charge states were manually assigned. Excel was used to generate zero-charge (deconvoluted) spectra by multiplying the m/z values (from Origin) by assigned charge states and subtracting appropriate protons. The spectra were input into ProsightPTM in its absolute mass mode using average masses, and a precursor mass window of 2000 Da and a fragment tolerance of 1.5 Da were used for all searching. Probability scores (p-score) are based on a Poisson distribution that reflects the probability of matching identified product ions to a given protein sequence; scores equal to or lower than 0.05 are statistically significant.[9, 10] Expectation values equal the probability score multiplied by the number of protein sequences in the *Bos taurus* database (i.e. expectation value = p-score x 2,869,593 protein

forms).[11] The PDE score, which is based on the McLuckey score,[12] takes into account product ion intensities and places weighting factors that increase scores when product ions stem from cleavages N-terminal to proline (5x weighting factor), C-terminal to aspartic acid (5x weighting factor), C-terminal to lysine (4x weighting factor), C-terminal to glutamic acid (2x weighting factor), N-terminal to proline and C-terminal to aspartic acid (10x weighting factor), N-terminal to proline and C-terminal to lysine (9x weighting factor), and N-terminal to proline and C-terminal to glutamic acid (7x weighting factor). These cleavages are particularly prominent as compared to other backbone locations. Higher PDE scores reflect better confidences in protein identification.

2.9.2 SEQUEST

Prior to SEQUEST interpretation of the MS/MS data, UVPD mass spectra were subjected to a background subtraction procedure to reduce/eliminate photoionization products. Background spectra were collected when no sample was being infused and thus no ions were detected in the trap, and at isolation *m/z* values of 250, 500, 750, or 1000 (isolation width of 4 *m/z*), using a protocol discussed in more detail later. Each centroid background spectra was the culmination of 20 averages of three microscans. The LC-MS/UVPD RAW files were subtracted from the background RAW files using Thermo Fisher Xcalibur version 2.0.7 software.

SEQUEST was used for *in silico* MS/MS interpretation through the Thermo Fisher Scientific Proteome Discoverer 1.0 software package. For SEQUEST, a signal:noise ratio of 3, a precursor mass tolerance of 1.5 Da, and a fragment mass tolerance of 0.8 Da were used. Product ion series for UVPD included *a*, *b*, *c*, *x*, *y*, and *z* ions, and for CID included *b* and *y* ions except where noted. UVPD and CID spectra were manually separated from LC-MS-UVPD/CID files for direct comparison. Non-redundant,

bovine (25,243 sequences) and human (33,819 sequences) protein databases from the NCBI were used for searching of BSA and HT-1080 samples, respectively. Oxidation of methionines and carbamidomethyl modification of cysteines were set as dynamic and static (respectively) side chain modifications. Peptide hits were filtered against decoy databases at a 1% false discovery rate (FDR) (based on Xcorr vs. charge), and were manually verified by the presence of matching immonium ions and predicted sequence ions (from selective cleavages). Peptides with less than six amino acids were also filtered out.

2.10 References

(1) Schwartz, J. C.; Syka, J. E. P.; Remes, P. M. *Proceedings of the 56th ASMS Conference on Mass Spectrometry and Allied Topics* **2008**, Denver, CO, June 2008.

(2) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80*, 4807-4819.

(3) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 6855-6862.

(4) Wang, D.; Kalb, S. R.; Cotter, R. J. *Rapid Communications in Mass Spectrometry* **2003**, *18*, 96-102.

(5) Shin, J.-W.; Lee, Y. H.; Hwang, S.; Lee, S.-W. *Journal of Mass Spectrometry* **2007**, *42*, 380-388.

(6) Taylor, G. K.; Kim, Y.-B.; Forbes, A. J.; Meng, F.; McCarthy, R.; Kelleher, N. L. *Analytical Chemistry* **2003**, *75*, 4081-4086.

(7) LeDuc, R. D.; Taylor, G. K.; Kim, Y.-B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. *Nucleic Acids Research* **2004**, *32*, W340-W345.

(8) Zamdborg, L.; LeDuc Richard, D.; Glowacz Kevin, J.; Kim, Y.-B.; Viswanathan, V.; Spaulding Ian, T.; Early Bryan, P.; Bluhm Eric, J.; Babai, S.; Kelleher Neil, L. *Nucleic Acids Research* **2007**, *35*, W701-706.

(9) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nature Biotechnology* **2001**, *19*, 952-957.

(10) Liu, J.; Huang, T.-Y.; McLuckey, S. A. *Analytical Chemistry* **2009**, *81*, 2159-2167.

(11) Sadygov, R. G.; Yates, J. R. *Analytical Chemistry* **2003**, *75*, 3792-3798.

(12) Reid, G. E.; Shang, H.; Hogan, J. M.; Lee, G. U.; McLuckey, S. A. *Journal of the American Chemical Society* **2002**, *124*, 7353-7362.

Chapter 3

Comparison of Infrared Multiphoton Dissociation and Collision Induced Dissociation of Supercharged Peptides in Ion Traps

3.1 Overview

The number and types of diagnostic ions obtained by infrared multiphoton dissociation (IRMPD) and collision-induced dissociation (CID) were evaluated for supercharged peptide ions created by electrospray ionization of solutions spiked with *m*-nitrobenzyl alcohol. IRMPD of supercharged peptide ions increased the sequence coverage compared to that obtained by CID for all charge states investigated. The number of diagnostic ions increased with the charge state for IRMPD; however, this trend was not consistent for CID as the supercharged ions did not always yield the greatest number of diagnostic ions. Significantly different fragmentation pathways were observed for the different charge states upon CID or IRMPD with the latter yielding far more immonium ions and often fewer uninformative ammonia, water, and phosphoric acid neutral losses. Pulsed-q dissociation (P-QD) resulted in an increase in the number of internal product ions, a decrease in sequence-informative ions, and reduced overall ion abundances. The enhanced sequence coverage afforded by IRMPD of supercharged ions was demonstrated for a variety of model peptides, as well as for a tryptic digest of cytochrome c.

3.2 Introduction

Collision-induced dissociation (CID) is the most common method used for sequencing peptides and identifying PTMs,[1] but uninformative labile neutral losses for

both unmodified and modified peptides remain a consistent problem. Recently, electron capture dissociation (ECD)[2] in FT-ICR instruments, a method that involves the reaction of multiply-charged cations with low energy electrons, and electron transfer dissociation (ETD)[3] in ion trap instruments, a technique which exploits the reaction of multiply-charged cations with radical anions, have been developed to combat this neutral loss problem by affording complementary *c*- and *z*-type backbone cleavages that leave labile modifications intact. These methods have been particularly promising for characterization of PTMs;[4-8] however, both techniques suffer from low fragmentation efficiencies, especially for lower charge states,[9, 10] as compared to CID.

Infrared multiphoton dissociation (IRMPD) is another tandem MS method that has been successfully implemented in FT-ICR, time-of-flight, and quadrupole ion trap instruments systems.[6, 11-24] Much of the appeal of IRMPD in quadrupole ion traps is related to the broader *m/z* trapping range possible than for conventional CID. This stems from the ability to reduce the rf trapping voltage during ion activation, an option that is detrimental for CID due to the decrease in energy deposition associated with lower rf trapping voltages.[25] This low-mass cutoff (LMCO) problem associated with CID prohibits the detection of many diagnostic *b* and *y* ions of lower *m/z* as well as immonium ions, the latter which are particularly useful in determining the amino acid composition of unknown peptides. Furthermore, these lower *m/z* diagnostic ions are important for de novo sequencing algorithms and can be useful for the rapid determination of modified amino acids (i.e. phosphorylated, acetylated, and methylated residues).[26-28] Because all ions are continuously irradiated during the activation period, IRMPD can also promote secondary dissociation of primary product ions and thus lead to formation of a more diverse array of diagnostic ions compared to CID.[25, 29] Several methods have been explored to combat the low IRMPD efficiencies of larger peptides. For example, the

Glish group developed thermally-assisted IRMPD (TA-IRMPD),[25] which involved heating the bath gas in the ion trap to increase the internal energies of ions. Our group has pursued peptide derivatization strategies to increase the photoabsorptivity of ions and/or decrease their critical energies for dissociation.[30-33]

Charge state has a significant impact on ion peptide dissociation due to changes in peptide conformation and proton mobility, with more highly charged states often yielding a greater array of structurally informative backbone cleavages.[9, 34] More highly-charged peptides also have increased coulombic repulsion which may facilitate dissociation compared to lower-charged species. Iavarone and Williams have explored the use of additives such as *m*-nitrobenzyl alcohol (*m*-NBA) that increase the surface tension and reduce the vapor pressure of ESI droplets, thus generating higher protein and peptide charge states.[35-39] This phenomenon has been termed supercharging. Tandem mass spectrometry has recently been employed for the analysis of these supercharged ions, revealing interesting fragmentation trends such as enhanced formation of diagnostic ions upon CID of the highest supercharged state of proteins[38] and increased sequence coverage upon ETD of the highest supercharged state of tryptic peptides.[9] Although targeting ions in higher charge states can potentially increase the number of redundant product ions (i.e., ones formed in multiple charges states), this has not proven to be a pervasive problem.

In the present study, the IRMPD and CID spectra of supercharged peptides obtained in ion trap mass spectrometers are compared. The more highly charged ions have lower critical energies; however, increased sequence coverage is most notable for IRMPD, which is attributed in part to the ongoing IR absorption and secondary dissociation of primary sequence ions throughout the activation period. Moreover, IRMPD allows detection of a greater array of diagnostic low *m/z* ions. Activation of

higher charge states is particularly advantageous for IRMPD since collisional cooling is competitive with energization due to the very low energy deposition per IR photon.[40] A comparison of P-QD to IRMPD and CID is also assessed in this study.

3.3 Experimental

3.3.1 Peptides, Proteins, and Reagents

All peptides, proteins, and reagents were purchased from Sigma-Aldrich (St. Louis, MO) except for the following: the peptides DRVYIHPFHLVIH and DAEFRHDSGYEVHHEK were purchased from BACHEM (King of Prussia, PA), the phosphorylated peptides KRpTIRR and NRVpYIHPF from AnaSpec (San Jose, Ca), and immobilized TPCK trypsin beads were purchased from Pierce Biotechnology, Inc. (Rockford, IL).

3.3.2 ESI-MS/MS

For conventional analysis, model peptides were diluted to approximately 10 μM with 49.5%/49.5%/1.0% (v/v) water/methanol/acetic acid solution prior to direct infusion at 3 μL/min into the mass spectrometer. Peptides were supercharged by the addition of 1 - 2% (v/v) *m*-NBA into the above working solutions prior to MS analysis.

Model peptides were analyzed using a modified Thermo Fisher (San Jose, CA) LTQ two-dimensional linear ion trap mass spectrometer equipped with a Synrad 50-W continuous wave CO_2 laser (model 48-5; Mukilteo, WA). Laser radiation at a wavelength of 10.6 μm was introduced into the linear ion trap via a modified back plate containing a ZnSe window.[33] Tryptic peptides were analyzed using a modified ThermoFinnigan LCQ Deca XP three-dimensional quadrupole ion trap mass spectrometer equipped with the

same Synrad 50-W continuous wave CO_2 laser.[30] Radiation at 10.6 μm was introduced into the quadrupole ion trap through a 5 mm hole drilled into the ring electrode. With the laser operated at full power, irradiation times of 10 – 250 ms (LTQ) and 25 – 500 ms (LCQ) were used with an activation q value of 0.1 during IRMPD experiments. For certain tryptic peptides, the He bath gas pressure was reduced during laser irradiation to allow for greater fragmentation efficiencies. A higher q value of 0.25 was used for CID experiments to afford more efficient ion activation at the expense of a more limited *m/z* trapping range. For all charge states, the CID voltage or IR irradiation time was adjusted to cause dissociation of approximately 100% (if possible) of the precursor ions. P-QD experiments were performed on the LTQ mass spectrometer with a q value of 0.7 and an activation time of 0.1 ms. For the present study, a product ion is considered to be detected in a CID, P-QD, or IRMPD mass spectrum if the ion peak has a S/N equal to or greater than three. Dissociation efficiencies were calculated based on the percentage of precursor ion abundance (based on ion peak area) converted into product ions, where a value of 100% signifies that the initial abundance of the isolated precursor ion is completely accounted for by the summed abundances of all product ions. Sequencing efficiencies (SEFF) were calculated based on the percentage of precursor ion abundance (based on ion peak area) converted into diagnostic ions (e.g. *b*, *y*, and immonium). A value of 100% indicates that the entire abundance of the selected precursor ion is converted into sequence-informative product ions. Although neutral losses, such as loss of ammonia or water, can give some useful information pertaining to the sequence of a peptide, these ions are often redundant or complicate spectra and are thus not classified as informative sequence ions. SEFF is a more useful parameter for assessing the diagnostic value of an MS/MS experiment, since product ions produced from neutral losses, such as ammonia loss or via dehydration, and internal fragments are not included. Moreover, redundant

multiply charged product ions were not counted in total diagnostic ion counts nor included in SEFF calculations. For the objectives of the present study, the term "immonium ion" is used to encompass both immonium and immonium-related ions.

3.3.3 Tryptic Digest

Cytochrome c (50 μM) was digested with 100 μL immobilized TPCK trypsin beads in 50 mM ammonium bicarbonate incubated for 18 hours at 37 °C. After digestion, the excess beads were discarded and the remaining tryptic peptides were diluted to 10 μM with 49.5%/49.5%/1.0% water/methanol/acetic acid solution prior to analysis. The tryptic peptides were supercharged by the addition of 1 - 2% *m*-NBA. Working solutions were directly infused into the mass spectrometer at 3 μL/min.

3.4 Results and Discussion

In this chapter, IRMPD experiments were undertaken by using both a two-dimensional linear ion trap mass spectrometer and a three-dimensional quadrupole ion trap mass spectrometer. IRMPD yielded the same number of diagnostic ions for both instruments; however, the irradiation times were approximately four times lower for the linear ion trap. The improved dissociation efficiency for the linear ion trap is attributed to a longer laser path length and/or better overlap of the laser with the ion cloud. For a more systematic investigation, the linear ion trap was used for all CID and IRMPD comparisons of model peptides, and the quadrupole ion trap was used for all CID and IRMPD comparisons of tryptic peptides.

3.4.1 Model Peptides

The addition of *m*-NBA to the peptide solutions typically increased the maximum observed charge state by one through the addition of one extra proton with little overall impact on total ion abundances. For example, by conventional ESI-MS analysis the peptide KFHEKHHSHRGY yielded a highest charge state of 4+ as seen in **Figure 3.1A**. However, with the addition of *m*-NBA the 5+ supercharged state emerged as the dominant ion (**Figure 3.1B**). All four observed charge states of the peptide were subjected to IRMPD as illustrated in **Figure 3.1C – 3.1F**. The IRMPD spectrum of the 2+ charge state is dominated by the intact precursor ion and only a few product ions of low abundance. The triply protonated peptide also exhibits relatively low IR photodissociation efficiency (i.e. only 52% of the original isolated precursor ion is converted into product ions), with much of the precursor ion surviving the irradiation period and only a modest number of diagnostic ions formed. Exposure of the 2+ and 3+ ions to long irradiation times (250 ms) did not yield a high degree of product ions, a factor attributed to the competition between IR energization and collisional cooling. However, the higher charge states (e.g. 4+ and 5+) required shorter irradiation times (100 ms), displayed much higher IRMPD efficiencies (up to 75% conversion of parent ions into total product ions), and yielded greater numbers of diagnostic ions. The sequencing efficiency (SEFF) ranged from 8% for the doubly charged peptide up to 39% for the highest charge state (5+). Activation of the highest charge state by IRMPD resulted in the greatest proportional yield of sequence-informative ions (*b*, *y*, and immonium). The 5+ supercharged state of KFHEKHHSHRGY also produced the most extensive sequence coverage upon IRMPD with six *b* ions, ten *y* ions, and three immonium ions. For CID of this same peptide (**Figure 3.2**), all charge states yielded a similar number of sequence ions (on average four *b* ions and seven *y* ions). The 3+ charge state yielded the greatest

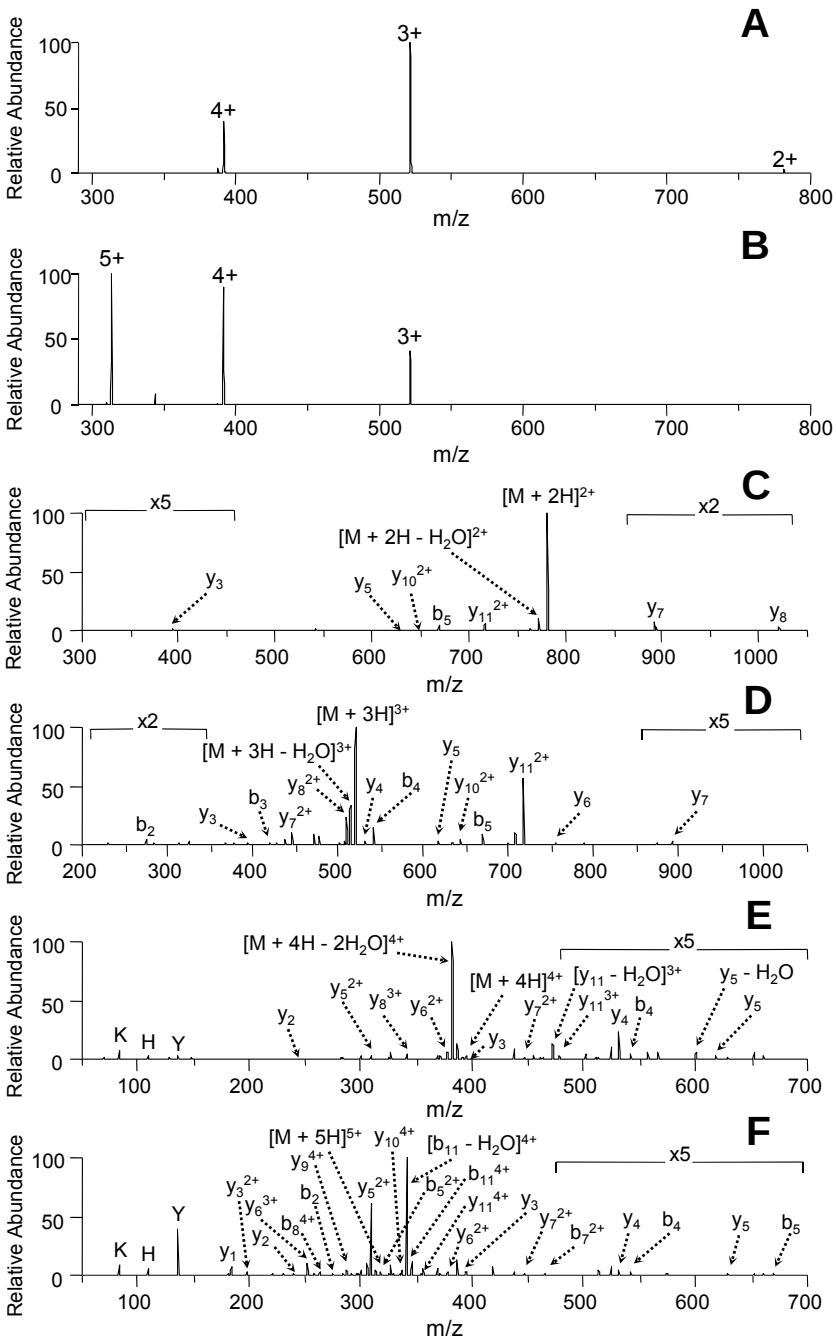

Figure 3.1 ESI mass spectra of KFHEKHHSHRGY (a) with 1% acetic acid and 49.5%/49.5% MeOH/H_2O (b) supercharged with 1% m-NBA, 1% acetic acid, and 49%/49% MeOH/H_2O. IRMPD mass spectra of the protonated peptide KFHEKHHSHRGY (c) 2+ charge state, 50 watts, t = 250 ms (d) 3+ charge state, 50 watts, t = 250 ms (e) 4+ charge state, 50 watts, t = 100 ms (f) 5+ supercharged state, 50 watts, t = 100 ms.

number of diagnostic ions upon CID, yet produced only 13 total sequence ions (six *b* ions and seven *y* ions) and no immonium ions. High CID dissociation efficiencies were achieved for all charge states, ranging from 68% to 87% conversion of the selected precursor ions into product ions. More importantly, SEFF values ranged from 16% for the 4+ charge state up to 36% for both the +3 and +5 charge states, all of which were lower than the supercharged state (5+) dissociated by IRMPD.

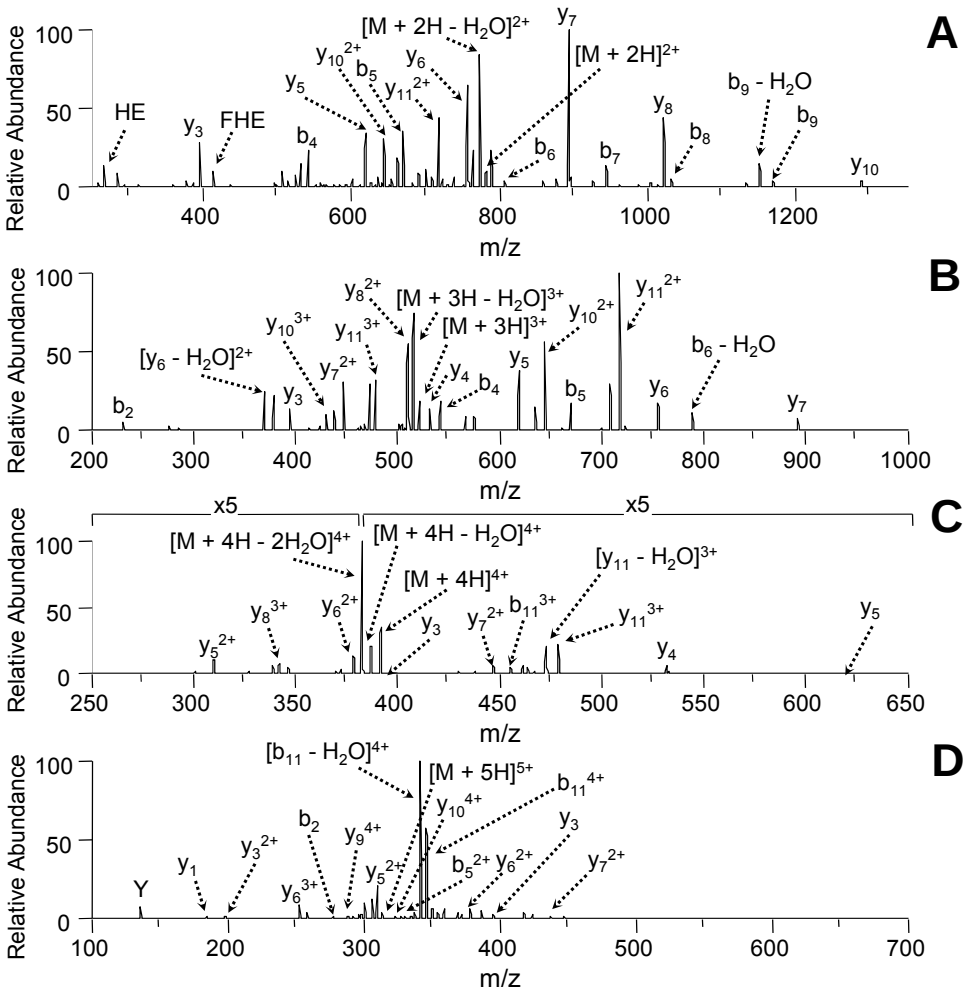

Figure 3.2 CID mass spectra of the protonated peptide KFHEKHHSHRGY (a) 2+ charge state, 29 mV collision voltage (b) 3+ charge state, 24 mV collision voltage (c) 4+ charge state, 17 mV collision voltage (d) 5+ supercharged state, 16 mV collision voltage.

The supercharged states also yielded greater sequence coverage for peptides that readily absorb IR radiation, such as phosphorylated peptides. As seen for the series of IRMPD spectra of the phosphorylated peptide KRpTIRR in **Figure 3.3**, the loss of phosphoric acid dominates the spectra for the lower charge states (1+ to 3+). In contrast,

the greatest array of diagnostic ions is seen for the 4+ peptide by IRMPD. **Figure 3.3** illustrates that the total sequence coverage increases with each additional charge. Phosphoric acid loss is significant upon IRMPD of the lower charge states but has less of an impact for the higher charge states, and is still less substantial than that observed upon CID.

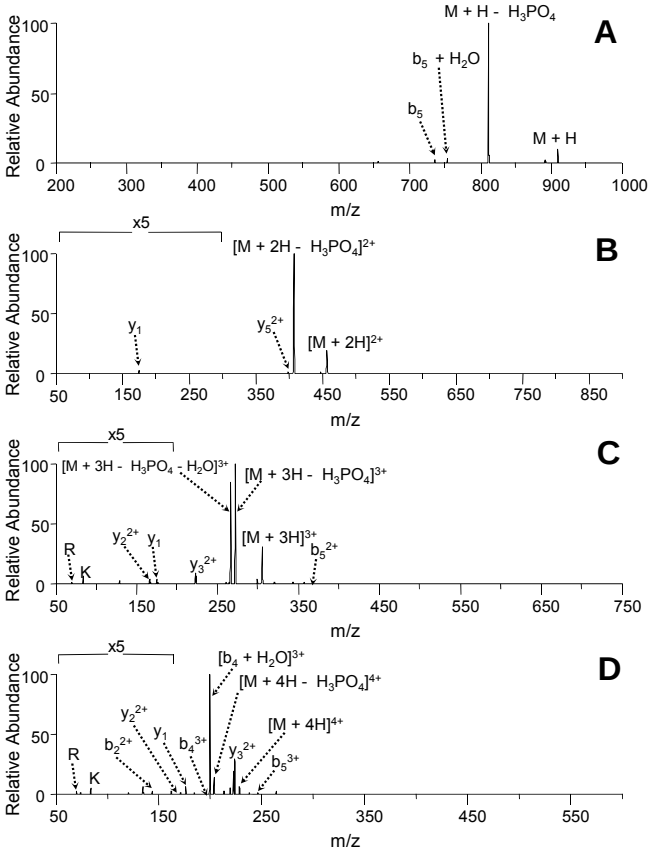

Figure 3.3 IRMPD mass spectra of the protonated phosphorylated peptide KRpTIRR, (a) 1+ charge state, 50 watts, t = 14 ms (b) 2+ charge state, 50 watts, t = 14 ms (c) 3+ charge state, 50 watts, t = 14 ms (d) 4+ supercharged state, 50 watts, t = 10 ms.

A series of eight peptides were analyzed by both IRMPD and CID to explore the impact of supercharging on the sequence coverage for peptides as a function of their size (6 to 16 amino acids) and phosphorylation status. The results are summarized in bar

graph form in **Figure 3.4**. For peptides subjected to IRMPD as seen in **Figure 3.4A**, the number of diagnostic ions increased with increasing charge state, with the supercharged state always producing the greatest number of informative ions. For CID (**Figure 3.4B**), however, this trend was not consistent and the supercharged ion did not always yield the greatest number of diagnostic ions. Activation of the higher charged ions often required lower CID voltages than lower charge states. Those peptides for which the higher charge states produced fewer diagnostic ions upon CID typically underwent dehydration which accounts for their less diagnostic fragmentation patterns. For IRMPD, however, fewer of these uninformative neutral losses were seen for the supercharged states which may be in part due to rapid secondary dissociation of these primary product ions into more useful diagnostic ions. This phenomenon could explain the general enhancement in the total number of informative product ions observed for IRMPD compared to CID.

Figure 3.4C shows a side-by-side comparison of CID and IRMPD for the best performing charge states (i.e. the charge state which produced the greatest total number of diagnostic ions) for each model peptide included in the study. This bar graph comparison confirms that IRMPD of the supercharged peptide uniformly affords the greatest sequence coverage. This increased sequence coverage for IRMPD is attributed to a combination of three factors: 1) reduction in the ion critical energies for higher charge states due to some combination of coulombic effects and enhanced proton mobility, 2) broader *m/z* trapping range due to alleviation of the low mass cutoff, and 3) ongoing IR absorption and secondary dissociation that leads to a greater array of ions and transformation of less informative ions (i.e. product ions from neutral loss of water, ammonia, or phosphoric acid loss) into more diagnostic ones.

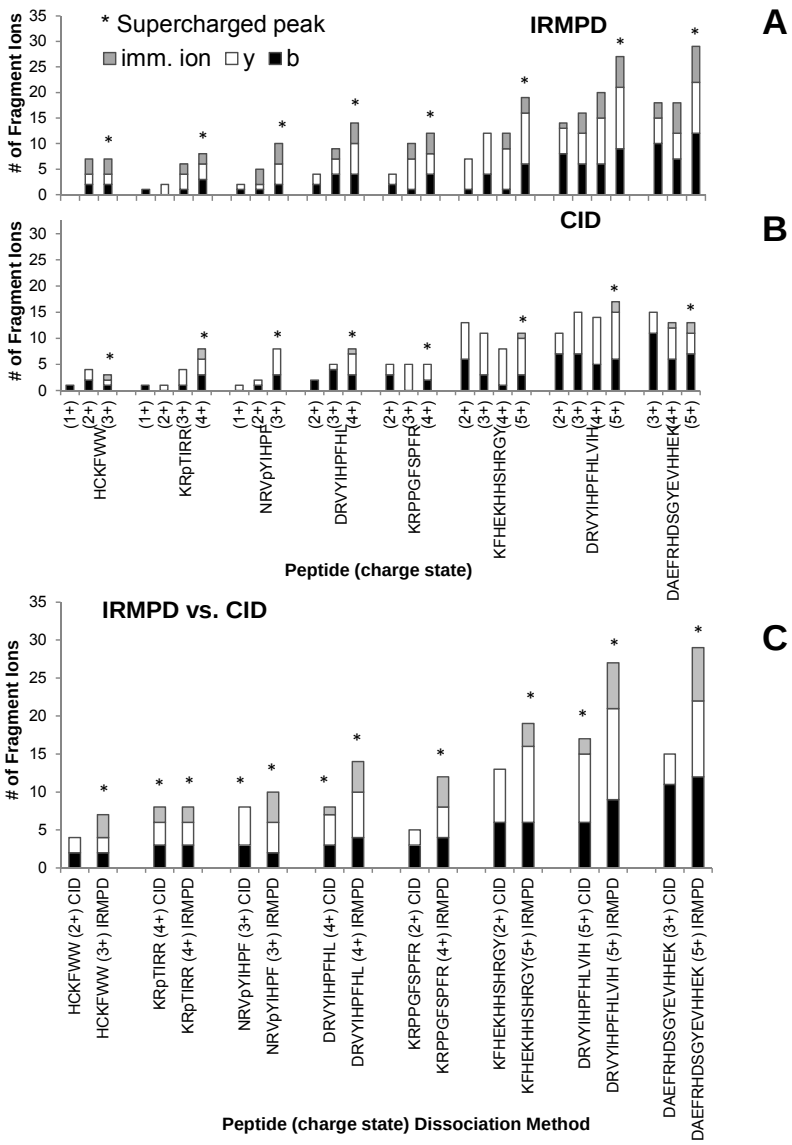

Figure 3.4 Number of diagnostic ions (*y*, *b*, and immonium) from different charge states of protonated peptides (a) using IRMPD (b) and CID. The best performers (i.e. most diagnostic ions) for each method (e.g. IRMPD and CID) are compared in (c). The average IRMPD irradiation times were 98 ms (1+), 116 ms (2+), 84 ms (3+), 51 ms (4+), and 48 ms (5+). Activation times for CID were 30 ms with average collisional voltages of 40 mV (1+), 26 mV (2+), 20 mV (3+), 16 mV (4+), and 14 mV (5+).

3.4.2 Tryptic Digest of Cytochrome C

In a manner similar to that described above for the model peptides, IRMPD and CID were used to evaluate a series of tryptic peptides from cytochrome c. The tryptic digest was mixed with 1% *m*-NBA, and then each charge state of each tryptic peptide was analyzed by CID and IRMPD. The number of diagnostic product ions (*b, y*, and immonium ions) were tabulated, and the results are summarized in **Figure 3.5A** and **3.5B**, respectively. An increasing number of diagnostic ions were formed for higher peptide charge states by IRMPD of the tryptic peptides, a trend not observed for CID. Also, IRMPD of the supercharged ions consistently yielded a greater number of diagnostic ions compared to CID. As confirmed again in a side-by-side bar graph that highlights the best CID results compared to the best IRMPD results for the series of tryptic peptides, IRMPD of each supercharged peptide afforded the greatest number of diagnostic ions in every case (**Figure 3.5C**). *m*-NBA has been used previously to create supercharged ions of tryptic peptides analyzed by LC-MS, [9] and we expect that this strategy combined with IRMPD would further facilitate protein characterization.

The dissociation efficiencies for conversion of precursor ions into product ions were also assessed for various tryptic peptides and charge states. For example, for the 1+ charge states of EDLIAYLK and GITWK, only 26% and 18% of the parent ions, respectively, were converted into product ions upon IR irradiation. Conversely, the IRMPD efficiencies increased to 43% and 48%, respectively, for the 2+ charge state. In comparison, the CID efficiencies ranged from 32% to 53% and 36% to 51% for the two charge states of EDLIAYLK and GITWK, respectively. For the peptide IFVQKCAQCHTVEK, the IRMPD efficiencies ranged from 55% for the 1+ charge state to 61% for the doubly-charged state to 57% for the triply-charged state. In comparison, the CID efficiencies ranged from 47% - 61% for the same peptide. Similar IRMPD

efficiencies and trends were seen for the 1+, 2+, and 3+ charge states of TGPNLHGLFGR and the 2+ and 3+ charge states of HKTGPNLHGLFGRK.

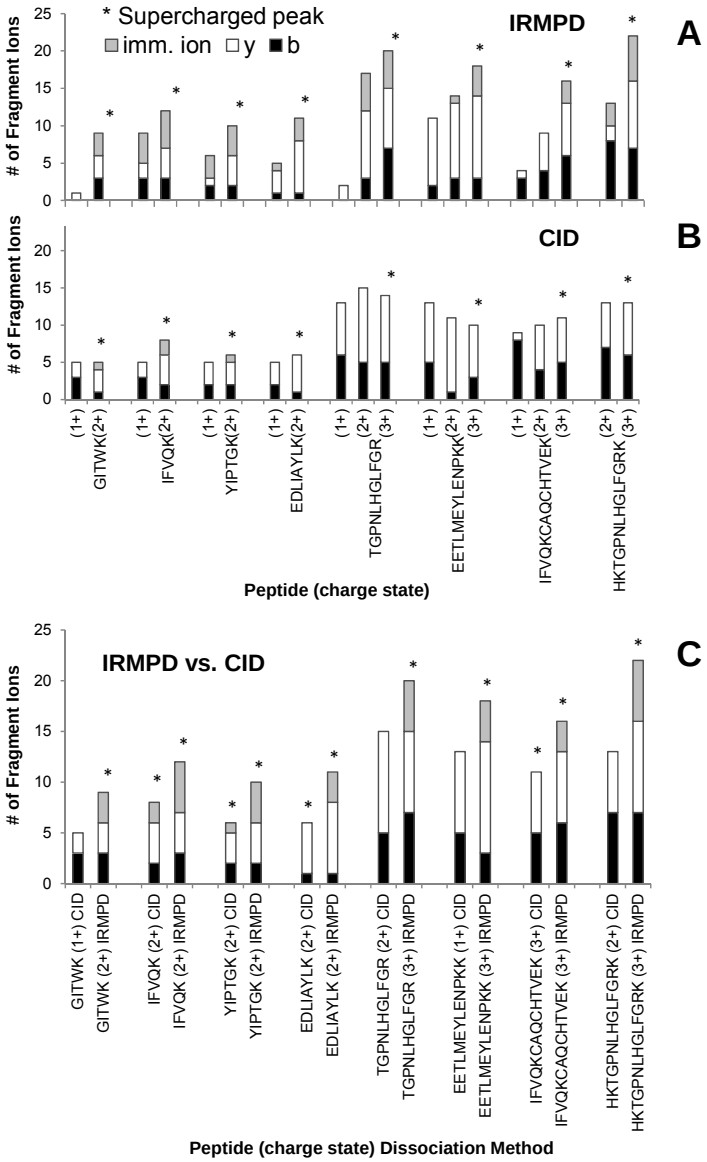

Figure 3.5 Number of diagnostic ions (*y*, *b*, and immonium) of various tryptic peptides from cytochrome c (a) using IRMPD (b) and CID. The best performers (i.e. most diagnostic ions) for each method (e.g. IRMPD and CID) are compared in (c). The average IRMPD irradiation times were 209 ms (1+), 123 ms (2+), and 50 ms (3+). Activation times for CID were 30 ms with average collisional voltages of 0.93 V (1+), 0.53 V (2+), and 0.42 V (3+).

SEFF values were also compared to the more traditional dissociation efficiencies for the tryptic peptides described above. CID of EDLIAYLK and GITWK yielded SEFF values ranging from 6% - 5% and 5% - 10% for the 1+ and 2+ charge states, respectively. Conversely, SEFF values obtained by IRMPD for the same two peptides in the singly-charged and doubly-charged states ranged from 4% - 27% and 2% - 30%, respectively. For the peptide IFVQKCAQCHTVEK, the IRMPD SEFF values ranged from 8% for the 1+ charge state to 19% for the 2+ charge state and up to 21% for the 3+ charge state. In comparison, the CID SEFF values ranged from 11% - 19% for the same peptide. Thus, although IRMPD often resulted in lower dissociation efficiencies (i.e. percentage of original isolated precursor converted into total product ions) as compared to CID, the SEFF values (original isolated precursor converted into diagnostic product ions) for the supercharged states were often higher for IRMPD. This increase in the SEFF values for IRMPD of the highest charge states reflects both the diminishment in uninformative neutral losses (e.g., H_2O, NH_3) and reduction in the formation of internal ions compared to CID.

Figure 3.6 shows the trends in the distributions of *b*, *y*, and immonium ions for one representative tryptic peptide, HKTGPNLHGLFGRK, in a quadrupole ion trap obtained by CID as the CID voltage is varied or by IRMPD as the irradiation time is varied. **Figures 3.6A** and **3.6B** show the distributions for the 3+ charge state, and the results for the 2+ charge state are shown in 6c and 6d. At longer irradiation times, IRMPD yielded significantly different distributions of product ions, including immonium ions and more *b* and *y* ions, compared to CID. For IRMPD, higher irradiation times were needed to dissociate the doubly-charged peptide, and the sequence coverage was lower. For CID, the total number of sequence ions observed was the same for both charge states. On average, a total of twelve *b* and *y* ions were seen for both charge states using CID,

with one additional *b* ion and one fewer *y* ion detected for the 2+ charge state as compared to the triply-charged species. For the +2 charge state, an average of seven *b*, two *y*, and three immonium ions were seen when using IRMPD. Conversely, the average total *b* and *y* ion count increased to sixteen (seven *b* and nine *y* ions), and an average of five immonium ions were seen with IRMPD of the +3 charge state.

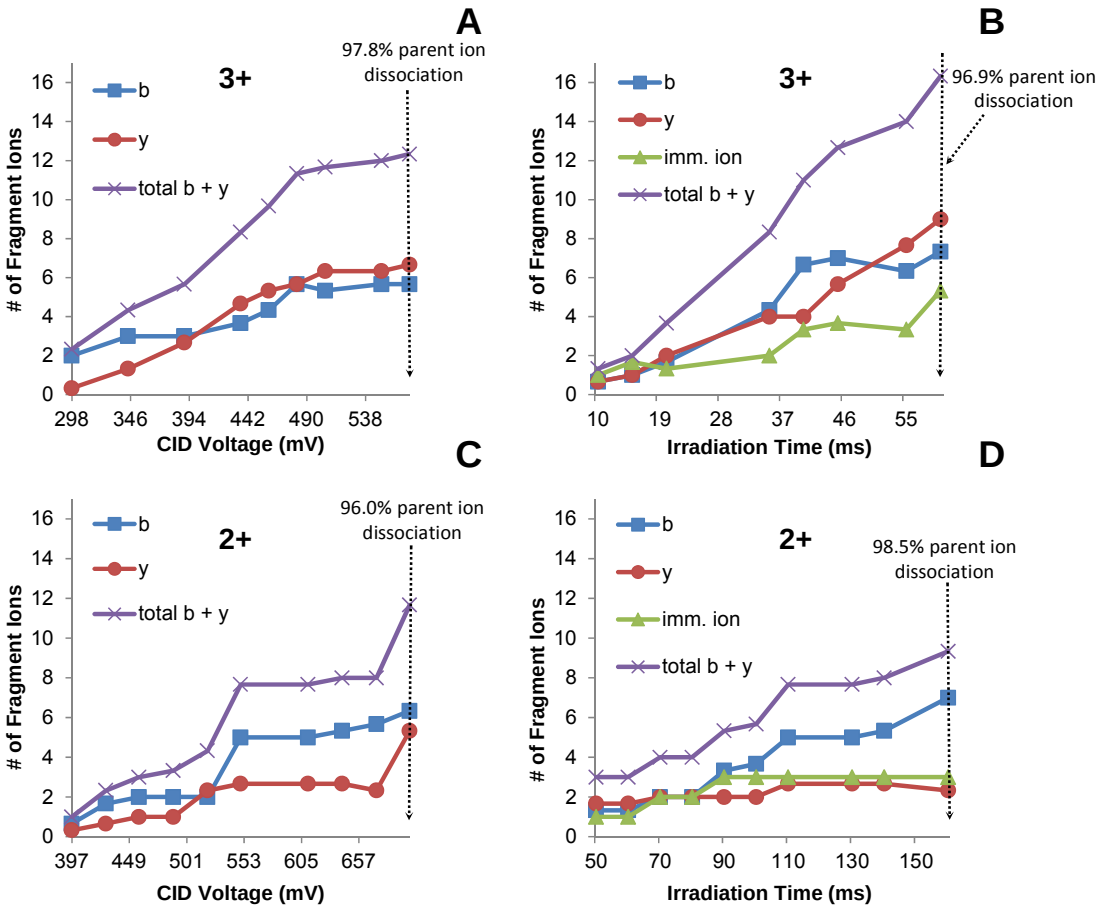

Figure 3.6 The distribution of diagnostic ions as a function of CID voltage or IRMPD irradiation time for protonated HKTGPNLHGLFGRK: (a) CID (3+), (b) IRMPD (3+), (c) CID (2+) and (d) IRMPD (2+).

3.4.3 Secondary Dissocation of Product Ion using MS^3

To further explore the occurrence of secondary dissociation during IRMPD, a product ion stemming from the neutral loss of water and another that incorporates the addition of water were selectively isolated and subjected to a second stage of ion activation in MS^3 experiments (**Figure 3.7**). The 3+ charge state of the peptide DRVYIHPFHL was collisionally activated to promote dehydration, resulting in the $[M-H_2O]^{3+}$ ion of high abundance. This product ion was then isolated and activated by IRMPD, thus resulting in the formation of two abundant *b* sequence ions and one immonium ion but very few secondary ions attributed to uninformative losses of water or ammonia. The triply-charged DRVYIHPFHL also produced a primary product ion, $[b_9 + H_2O]^{2+}$, of high abundance upon CID. This product ion was isolated and subjected to IRMPD, yielding three abundant *b* sequence ions and one immonium ion as well as a few internal product ions of low abundance. These MS^3 results suggest that the less informative neutral loss and neutral addition products ions that are commonly formed upon CID (or IRMPD) can be converted to more diagnostic product ions upon further activation. Since CID is a resonant process that causes energization of solely the selected precursor ion, it is only the non-resonant IRMPD method that can successfully convert the less informative primary product ions into more informative secondary product ions.

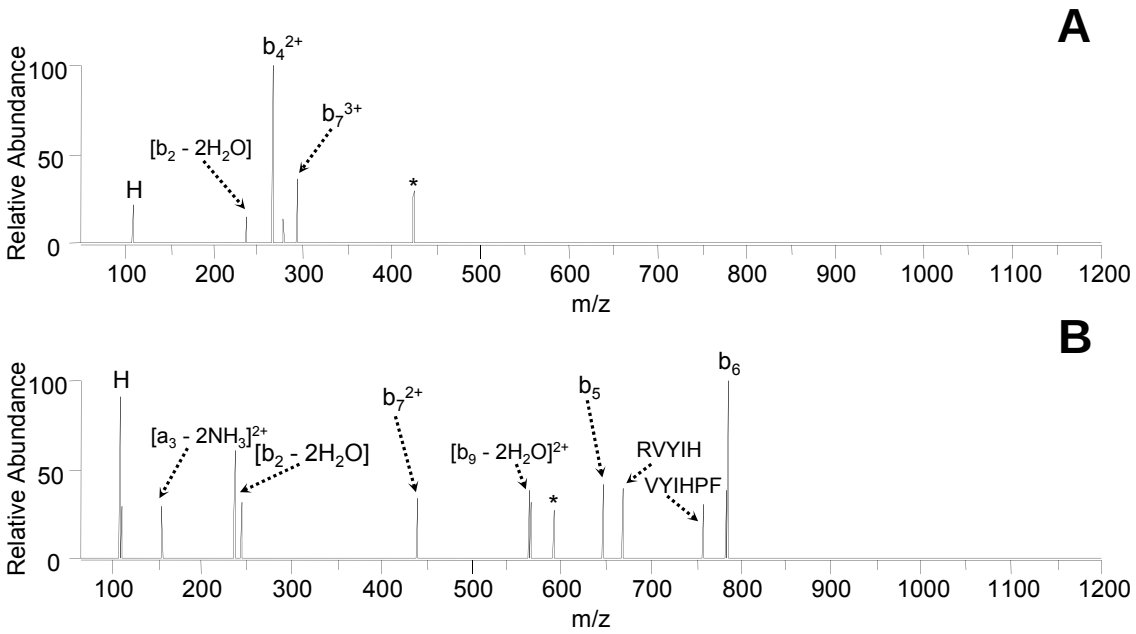

Figure 3.7 MS^3 spectra of (a) the isolated $[M+3H-H_2O]^{3+}$ ion and (b) the isolated $[b_9+H_2O]^{2+}$ ion produced upon CID and then dissociated by IRMPD at 50 watts and 75 ms from the peptide DRVYIHPFHL. The selected precursor ion is denoted by *.

3.4.4 P-QD Comparison

As a final comparison, the tryptic peptide HKTGPNLHGLFGRK in the 3+ charge state was subjected to P-QD[41] to assess the total number of diagnostic ions since this dissociation method, like IRMPD, alleviates the LMCO problem associated with CID. P-QD entails activating the precursor ion at a high q value and high CID energy for a short time, and then quickly lowering the trapping rf voltage to detect product ions of low *m/z*. At a P-QD voltage of 46 mV, only four diagnostic ions were observed with much of the ion signal attributed to undesirable internal fragments at m/z values that clutter the range where *b* and *y* ions would be found (**Figure 3.8**). As the P-QD voltage was increased to 53 mV, the total ion abundance was reduced by 85%, the level of noise increased, and a total of only two sequence ions were detected (spectrum not shown). Increasing the P-

QD voltage beyond 66 mV decreased the total ion abundance by over 97%, and the noise level obscured any fragment ion peaks. For comparison, when this triply-protonated peptide was subjected to IRMPD, twenty-two diagnostic ions were produced with a low abundance of uninformative internal fragment ions (**Figure 3.5A**).

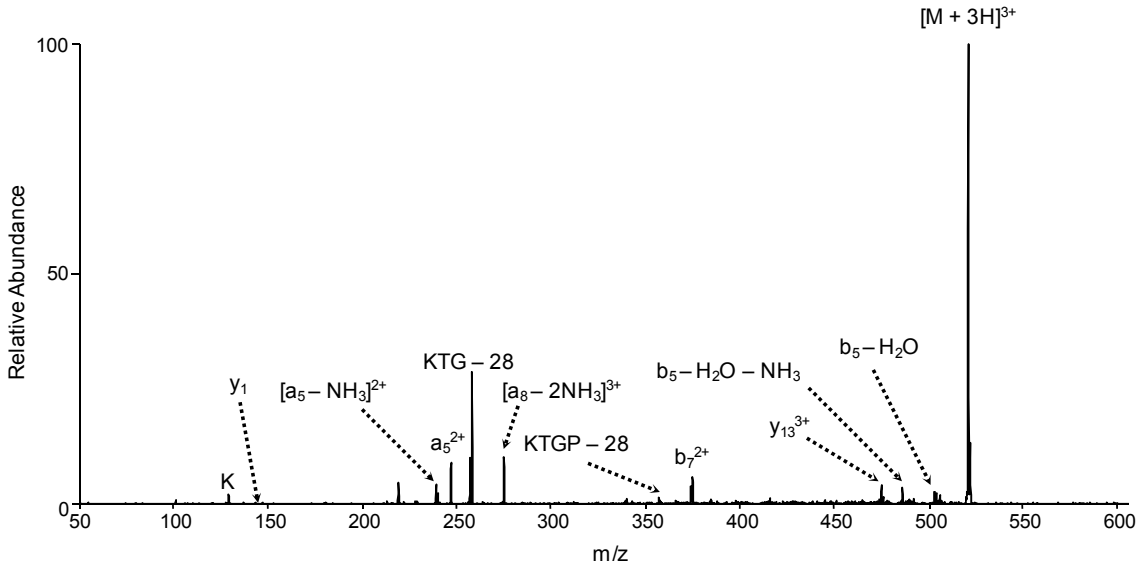

Figure 3.8 P-QD mass spectrum of triply protonated HKTGPNLHGLFGRK at 46 mV, q = 0.7, and an activation time of 0.1 ms.

3.5 Conclusions

IRMPD of the supercharged states of peptide ions resulted in the most informative array of sequence ions when compared to dissociation of lower charge states by IRMPD or upon dissociation of any charge state by CID. In addition, the number of immonium ions was dramatically increased with IRMPD of supercharged ions, which could aid in identification of PTMs. In general, the total number of diagnostic ions increased with increasing charge state for IRMPD. This trend was not observed for CID, and often lower charge states led to the greatest number of diagnostic ions. IRMPD of supercharged ions is anticipated to be particularly beneficial for larger peptides or proteins to maximize sequence coverage.

3.6 References

(1) Wells, J. M.; McLuckey, S. A. *Methods in Enzymology* **2005**, *402*, 148-185.
(2) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.
(3) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(4) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Analytical Chemistry* **1999**, *71*, 4431-4436.
(5) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 1793-1800.
(6) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Analytical Chemistry* **2001**, *73*, 4530-4536.
(7) Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 7450-7456.
(8) Srikanth, R.; Wilson, J.; Bridgewater, J. D.; Numbers, J. R.; Lim, J.; Olbris, M. R.; Kettani, A.; Vachet, R. W. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1499-1506.
(9) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 9243-9252.
(10) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Analytical Chemistry* **2007**, *79*, 477-485.
(11) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Analytical Chemistry* **1994**, *66*, 2809-2815.
(12) Li, W.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. *Analytical Chemistry* **1999**, *71*, 4397-4402.
(13) Hofstadler, S. A.; Sannes-Lowery, K. A.; Griffey, R. H. *Analytical Chemistry* **1999**, *71*, 2067-2070.
(14) Flora, J. W.; Muddiman, D. C. *Analytical Chemistry* **2001**, *73*, 3305-3311.
(15) Flora, J. W.; Muddiman, D. C. *Journal of the American Chemical Society* **2002**, *124*, 6546-6547.
(16) Drader, J. J.; Hannis, J. C.; Hofstadler, S. A. *Analytical Chemistry* **2003**, *75*, 3669-3674.
(17) Hofstadler, S. A.; Drader, J. J.; Gaus, H.; Hannis, J. C.; Sannes-Lowery, K. A. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 1413-1423.
(18) Chalmers, M. J.; Quinn, J. P.; Blakney, G. T.; Emmett, M. R.; Mischak, H.; Gaskell, S. J.; Marshall, A. G. *Journal of Proteome Research* **2003**, *2*, 373-382.
(19) Flora, J. W.; Muddiman, D. C. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 121-127.

(20) Chalmers, M. J.; Hakansson, K.; Johnson, R.; Smith, R.; Shen, J.; Emmett, M. R.; Marshall, A. G. *Proteomics* **2004**, *4*, 970-981.
(21) Tsybin, Y. O.; Ramstroem, M.; Witt, M.; Baykut, G.; Hakansson, P. *Journal of Mass Spectrometry* **2004**, *39*, 719-729.
(22) Yamada, N.; Suzuki, E.-I.; Hirayama, K. *Analytical Biochemistry* **2006**, *348*, 139-147.
(23) Mihalca, R.; van der Burgt, Y. E. M.; McDonnell, L. A.; Duursma, M.; Cerjak, I.; Heck, A. J. R.; Heeren, R. M. A. *Rapid Communications in Mass Spectrometry* **2006**, *20*, 1838-1844.
(24) Clowers, B. H.; Dodds, E. D.; Seipert, R. R.; Lebrilla, C. B. *Journal of Proteome Research* **2007**, *6*, 4032-4040.
(25) Payne, A. H.; Glish, G. L. *Analytical Chemistry* **2001**, *73*, 3542-3548.
(26) Steen, H.; Fernandez, M.; Ghaffari, S.; Pandey, A.; Mann, M. *Molecular & cellular proteomics : MCP* **2003**, *2*, 138-145.
(27) Trelle, M. B.; Jensen, O. N. *Analytical Chemistry* **2008**, *80*, 3422-3430.
(28) Couttas, T. A.; Raftery, M. J.; Bernardini, G.; Wilkins, M. R. *Journal of Proteome Research* **2008**, *7*, 2632-2641.
(29) Crowe, M. C.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 1581-1592.
(30) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 6855-6862.
(31) Pikulski, M.; Wilson, J. J.; Aguilar, A.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 8512-8517.
(32) Pikulski, M.; Hargrove, A.; Shabbir, S. H.; Anslyn, E. V.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 2094-2106.
(33) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80*, 4807-4819.
(34) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *Journal of Mass Spectrometry* **2000**, *35*, 1399-1406.
(35) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 976-985.
(36) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. *Analytical Chemistry* **2001**, *73*, 1455-1460.
(37) Iavarone, A. T.; Williams, E. R. *International Journal of Mass Spectrometry* **2002**, *219*, 63-72.
(38) Iavarone, A. T.; Williams, E. R. *Analytical Chemistry* **2003**, *75*, 4525-4533.
(39) Iavarone, A. T.; Williams, E. R. *Journal of the American Chemical Society* **2003**, *125*, 2319-2327.
(40) Black, D. M.; Payne, A. H.; Glish, G. L. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 932-938.
(41) Schwartz, J. C.; (Thermo Finnigan Llc, USA). Application: US, 2005, pp 11 pp.

Chapter 4

Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation

4.1 Overview

N-terminal peptide derivatization strategies used in conjunction with tandem mass spectrometry to yield simplified fragmentation patterns have shown limited success for the *de novo* sequencing of multiply charged peptides, including those predominantly formed in LC-ESI-MS experiments. Significant proton mobilization occurs for multiply charged peptides upon collisional activation, resulting in the formation of both N-terminal and C-terminal product ions rather than an exclusive series of C-terminal ions preferred for *de novo* sequencing algorithms. To circumvent this problem, multiply charged, N-terminally derivatized peptides were subjected to electron transfer reactions with fluoranthene anions to produce singly charged, radical species. Upon subsequent "soft" collision induced dissociation (CID), highly abundant *z*-type ions were formed nearly exclusively, which yielded simplified fragmentation patterns amenable to *de novo* sequencing methods. Furthermore, the simplified series of *z* ions were shown to retain labile phosphoric acid moieties and were also observed for peptides not possessing basic C-terminal residues – a type of peptide that poses more challenges to traditional collision induced dissociation. This improved LC-MS^n strategy was demonstrated for a variety of multiply charged model peptides and a tryptic digest of myoglobin.

4.2 Introduction

Mass spectrometry (MS) has increasingly become the analytical method of choice in the field of proteomics in recent years due to significant advances in instrumentation, sampling techniques, and data interpretation algorithms. Although there have been tremendous inroads, there also remain some unresolved limitations in practical applications. For example in a recent study conducted this year, 24 proteomics labs were given a sample of *Escherichia coli* spiked with 20 proteins and of these participating labs, only six correctly identified the 20 proteins.[1] While discouraging, these results were a significant improvement compared to a similar study conducted the previous year. One key problem resulting in protein misidentification is a high false positive rate in computer-guided interpretation of tandem mass spectra. In a typical "bottom-up" approach, proteins are first enzymatically digested, and then tandem MS is performed on the resulting peptides. *De novo* algorithms or cross-correlation scoring methods can then be used to decipher the tandem mass spectra, assign peptide sequences, and integrate the data sets to identify proteins.[2-8] However, often the tandem mass spectra are extremely complex with many redundant product ions that cause errors in spectral interpretation,[9-11] leading to incorrect identification of proteins or unidentified proteins. To enhance the success of *de novo* sequencing methods, measures have been taken to reduce the complexity of product ion spectra. Keough et al. first reported the derivatization of the N-terminus of peptides with sulfonic acid groups, which upon dissociation yield mass spectra with an extensive array of C-terminal ions (e.g. *y* ions) void of neutralized N-terminal ions (e.g. *b* ions).[12] This technique greatly improved the accuracy of *de novo* peptide sequencing methods. A number of other studies have described the potential merits of other derivatization procedures for proteomic applications.[10, 11, 13-34] For example, the N-terminal sulfonation reagent 4-sulfophenyl isothiocyanate (SPITC) in

conjunction with ESI-MS/MS and MALDI-MS/MS has proved to be useful identifying sites of ubiquitination.[18, 19] Our group has used N-terminal sulfonation with infrared multiphoton dissociation (IRMPD) to improve *de novo* sequencing of peptides by eliminating the low mass cutoff inherent to ion traps.[11] Recently, we have synthesized a new N-terminal reagent with a highly IR-absorbing phosphonite group that increased the dissociation efficiencies and sequence information obtained upon IRMPD.[35]

One major drawback to the use of N-terminal derivatization strategies is that the simplification of the fragmentation patterns of peptides is substantially impaired when multiply charged precursor ions are analyzed.[14, 15, 18, 20] For example, when doubly charged SPITC-modified peptides are dissociated, the N-terminal *b* ions are no longer neutralized due to the presence of the extra mobile proton, and the utility of the derivatization procedure is greatly reduced.[14] This is particularly problematic for LC-ESI-MS applications in which tryptic peptides are typically multiply protonated (i.e. charge states of 2+ and greater).[36] Lee et al. described a method to simplify the fragmentation patterns of these multiply charged ions by using isotopically-labelled SPITC, but extra derivatization steps and more elaborate spectral interpretation were needed.[15] Our group has attached UV chromophores to the N-terminus of peptides, and the modified peptides almost exclusively produced *y* ions upon ultra-violet photodissociation (UVPD), a striking selectivity attributed to the secondary dissociation and rapid annihilation of the UV chromophore-containing *b* ions.[34] However, due to the slow repetition rate of the laser (10 Hz) and the requirement for multiple laser pulses used in this application, analysis on a chromatographic timescale could not be achieved.

In recent years, other ion activation methods have been developed as alternatives to collision induced dissociation. Electron-based dissociation techniques such as electron capture dissociation (ECD)[37] in FT-ICR instruments and electron transfer dissociation

(ETD)[38] in ion traps have shown to be particularly promising for characterization of post-translational modifications (PTMs)[38-45] and for increasing peptide sequence coverage.[36, 46] Both methods involve the reaction of multiply charged cations with either low energy electrons (ECD) or radical anions (ETD), producing complementary *c*- and *z*-type backbone cleavages that retain labile PTMs such as phosphorylation. For peptides of lower charge, however, these electron-based methods yield charge-reduced radicals from the precursor as the most abundant products with low yields of *c* and *z* products.[36, 47] Recently, both the Coon and McLuckey groups have implemented supplemental collisional activation (i.e. low-energy CID) of charge-reduced peptide radicals after the electron transfer reaction (ETcaD), thus affording significantly improved peptide sequence coverage.[47-49] Both ETD and ECD are promising new alternatives to CID for sequencing peptides, but the resulting spectra remain rather complex with both N-terminus and C-terminus product ions that can complicate *de novo* interpretation.

In this study, doubly protonated peptides derivatized at the N-terminus via reactions with 4-sulfophenyl isothiocyanate (SPITC) or 4-(chlorosulfonyl)phenyl isocyanate (SPC) (**Figure 4.1**) were converted to charged-reduced radical species after gas-phase electron transfer (ET) reactions with fluoranthene radicals. The resulting singly-charged radical species were then subjected to subsequent collisional activation (ETcaD), which produced an enhanced series of *z* ions void of N-terminal ions. This spectral simplification strategy was performed using a linear ion trap on a liquid chromatographic timescale, and improved *de novo* sequencing was realized for a variety of multiply charged peptides including phosphorylated species.

4.3 Experimental

4.3.1 Materials

Myoglobin and cytochrome c from equine heart, and all reagents were purchased from Sigma-Aldrich (St. Louis, MO). The peptides KRPPGFSPFR and ASHLGLAR were obtained from BACHEM (King of Prussia, PA), the phosphorylated peptides KRpTIRR and TRDIYETDYpYRK from AnaSpec (San Jose, Ca), and immobilized TPCK-treated trypsin beads from Pierce Biotechnology, Inc. (Rockford, IL). All solvents were purchased from Fisher Scientific (Fairlawn, NJ).

4.3.2 Derivatization and Sample Preparation

Before N-terminal derivatization, lysines were converted to homoarginine residues through a guanidination step (except where noted) to ensure modification with 4-sulfophenyl isothiocyanate (SPTIC) exclusively at the N-terminus and not at the ε-amine group of lysines. A stock of 0.05 g O-methylisourea in 50 μL water was prepared and 5 μL of this stock solution, 50 μL of 5 N ammonium hydroxide, and approximately 10 nmol of the protein digest or model peptide was combined. This mixture was then incubated for 10 minutes at 65 °C, after which the sample was desalted via C_{18} spin columns.

The SPITC and 4-(chlorosulfonyl)phenyl isocyanate (SPC) derivatization procedures used in this study were adapted from previous reports.[16, 50] Briefly, a stock solution of SPITC was made by dissolving approximately 1 mg of SPTIC in 100 μL of a 20 mM $NaHCO_3$ solution (pH ~9.5). A 20 μL aliquot of this stock was combined with 10 nmol of peptide or protein digest, which was then incubated for approximately 30 minutes at 55 °C. For SPC-modified peptides, a stock solution was made by dissolving

approximately 1 mg of SPC in 100 μL acetonitrile. A 20 μL aliquot of this stock was combined with 10 nmol of peptide in a 50:50 pyridine/water (v/v) solution, which was then incubated for approximately 30 minutes at ambient temperature. After incubation, all samples were cleaned up using C_{18} spin columns. For direct infusion experiments, working solutions of 10 μM peptides in 49.5:49.5:1.0 water/methanol/acetic acid (v/v/v) solution were prepared before infusion at a flow rate of 3 μL/min into the mass spectrometer.

For protein samples, 100 μL of immobilized TPCK-treated trypsin beads and 10 micromoles of ammonium bicarbonate were used to digest 10 nmol of protein prior to the derivatization procedures. Digestion occurred for 18 hours at 37 °C, and the beads were discarded after digestion.

4.3.3 Mass Spectrometry and Liquid Chromatography

Mass spectrometric analysis was performed on a Thermo Fisher (San Jose, CA) LTQ XL two-dimensional linear ion trap mass spectrometer equipped with an ETD unit. Fluoranthene anions were introduced as the electron-transfer reagent for ETD experiments with reaction times of 100 ms. Standard parameters (q-value equal to 0.25 and activation times of 30 ms) were used for all CID experiments. ETcaD was carried out using ET reaction times of 100 ms followed by isolation of the charge-reduced radical, and a 30 ms CID period. For all MS^n experiments, a product ion was considered to be detected in an ETD, CID, or ETcaD mass spectrum if the ion peak had a signal: noise ratio equal to or greater than three.

A Hitachi L-7000 (Hitachi Ltd.) analytical HPLC system was used for liquid chromatographic separations. Samples containing 10 μM digested protein were injected (10 μL) onto a Symmetry300 reversed-phase C_{18} column (Waters, Milford, MA) (2.1 ×

50 mm, 3.5 μm packing) with a matched guard column (2.1 × 10 mm, 3.5 μm packing). Eluents consisted of 0.2% formic acid in water (A), and 0.2% formic acid in acetonitrile (B). Gradient elution was performed as follows: 95% (A) for 2 minutes followed by a linear gradient to 40% (A) over 60 minutes at a flow rate of 0.300 mL/min. Data-dependent aquisition was used for all LC-MS^n analysis with the first scan event being a full mass spectrum at a *m/z* range of 400 – 2000, and subsequent events consisting of ETD, CID, and ETcaD of the one or two most abundant peaks. A normalized collision energy of 35% (61 – 142 mV) was used for all LC-CID-MS analysis.

4.4 Results and Discussion

4.4.1 N-terminally Derivatized Multiply Charged Peptides

Two commonly used commercial reagents (e.g. SPITC and SPC) were used for N-terminal derivatization reactions of the peptides, resulting in a mass addition of 215 Da for SPITC and 199 Da for SPC. The structures of both reagents - SPITC and SPC - are shown in **Figure 4.1**. As noted previously, the chlorine atom in SPC is hydrolyzed and replaced with a hydroxyl group in aqueous solution.[16] The sulfonic acid group present in both reagents introduces a negative charge at the N-terminus, thus reducing the overall charge of the peptides and simplifying the tandem mass spectra by effectively neutralizing N-terminal product ions (e.g. *b*, *a*, and *c* ions). This simplification of the product ion spectra of singly charged sulfonate-derivatized peptides has been documented previously and has proven to be especially beneficial for *de novo* sequencing strategies.[10-12, 26] Upon CID, an enhanced set of C-terminal ions (*y* ions) are produced, and redundant N-terminal products (*b, a* ions) are formed as neutral species and are not detected.

However, when more than one positive charge resides on the derivatized peptide, the N-terminal products formed upon CID may retain a charge, thus nulling the benefit of the derivatization process for spectral simplification. This increased spectral convolution is illustrated in **Figure 4.2** for the doubly charged derivatized peptide ASHLGLAR in which N- and C-terminal product ions of multiple charge states are detected. Upon ESI, the N-terminal derivatized peptide ASHLGLAR was predominantly observed in the 2+ charge state (data not shown). CID of the doubly charged ions yielded complex spectra containing both *y* and redundant *a* and *b* ions (with these N-terminal ions retaining the original SPITC- or SPC-modification) (see **Figures 4.2A** and **4.2B**). Moreover, product ions in multiple charge states are observed, as is the case for the y_7 ion, which further complicates spectral interpretation. This loss in spectral simplification for multiply charged peptides is the primary reason that N-terminal sulfonation reactions have mainly been employed for MALDI-MS, not ESI-MS, applications.

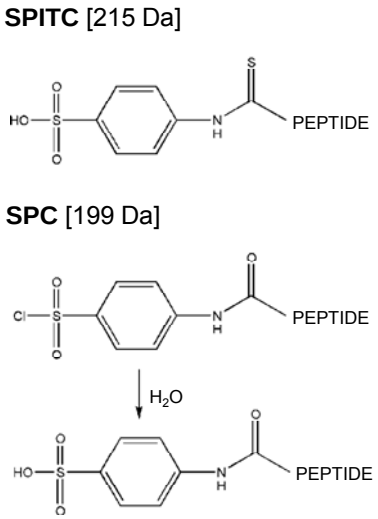

Figure 4.1 Structures of the N-terminal derivatization reagents 4-sulfophenyl isothiocyanate (SPITC) and 4-(chlorosulfonyl)phenyl isocyanate (SPC). The abbreviation and nominal mass additions are shown above the structures.

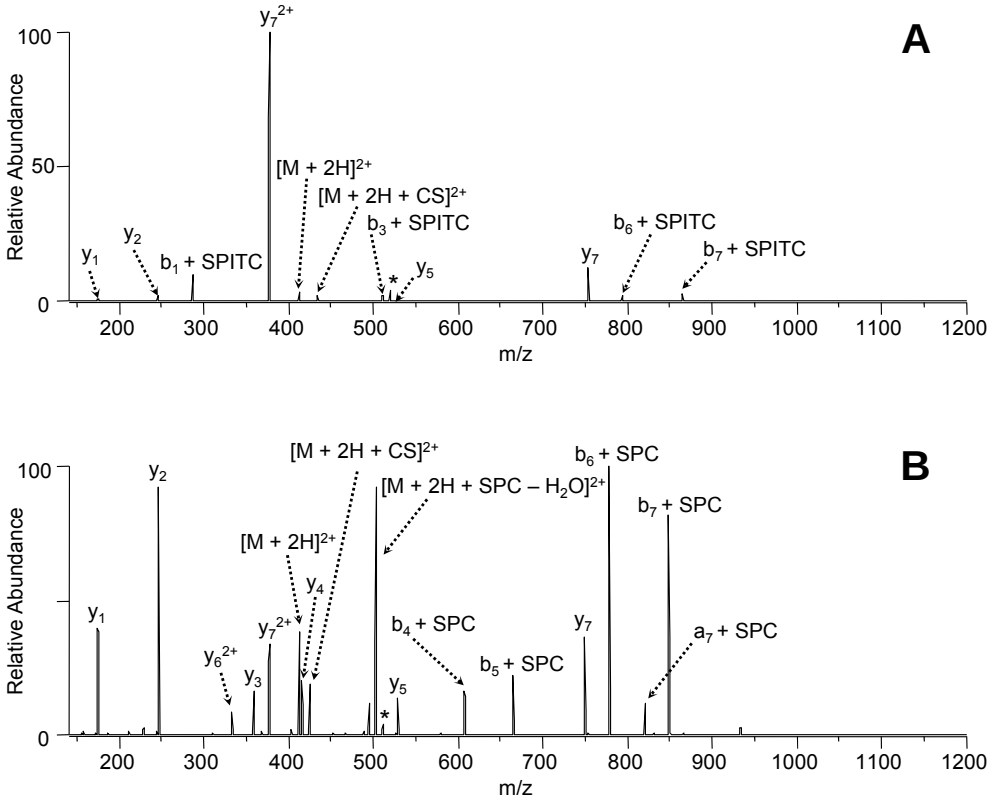

Figure 4.2 CID mass spectra of the doubly charged peptide ASHLGLAR (a) derivatized with SPITC, 12 mV collision voltage, q = 0.25; (b) derivatized with SPC, 14 mV collision voltage, q = 0.25. The unmodified peptide is denoted by M and the precursor ion by an asterisk (*). The partial decomposition of the SPITC or SPC moiety resulting in a bonded carbon and sulfur atom attached N-terminally to the peptide is denoted by CS.

An alternative approach, as described in the present study, relies on using ETD (via ETcaD) rather than CID for characterization of the N-terminal derivatized peptides. In this case, doubly charged SPITC- or SPC-modified peptides were subjected to electron transfer reactions, producing low amounts of sequence ions and predominantly charge-reduced peptide radical cations (singly charged) as illustrated in **Figure 4.3**. Subsequently, the resulting charged-reduced ions were subjected to collisional activation,

a net process termed ETcaD, which yielded a simple set of C-terminal (*z* ions) without any redundant N-terminal ions (*c, b,* or *a* ions) (see **Figure 4.4**). The ETcaD spectrum of doubly charged SPITC-modified ASHLGLAR is illustrated in **Figure 4.4A**, displaying a complete series of *z*-type ions and no N-terminal ions. Some *z* product ions were observed as radicals and others as even electron ions, a result which is discussed in more detail later. Also, a product ion attributed to the loss of both the SPITC moiety and ammonia (net loss of 232 Da) was observed in all ETcaD spectra; this characteristic loss can be used as a marker for every SPITC-modified peptide. It is a spectral feature that is especially convenient when the derivatization of peptide mixtures might be incomplete, and one needs a facile means to screen the species of interest. As also seen in **Figure 4.4A**, product ions attributed to partial cleavage of the SPITC moiety are observed in the *m/z* range of 850 to 1040. These product ions are thought to stem from the low energy, radical-driven dissociation processes of ETcaD since many of these ions were not observed upon CID of the singly-charged, even-electron peptides. Regardless, these ions do not cause substantial spectral congestion as they do not overlap with the *z* ion series and these were formed in low abundance.

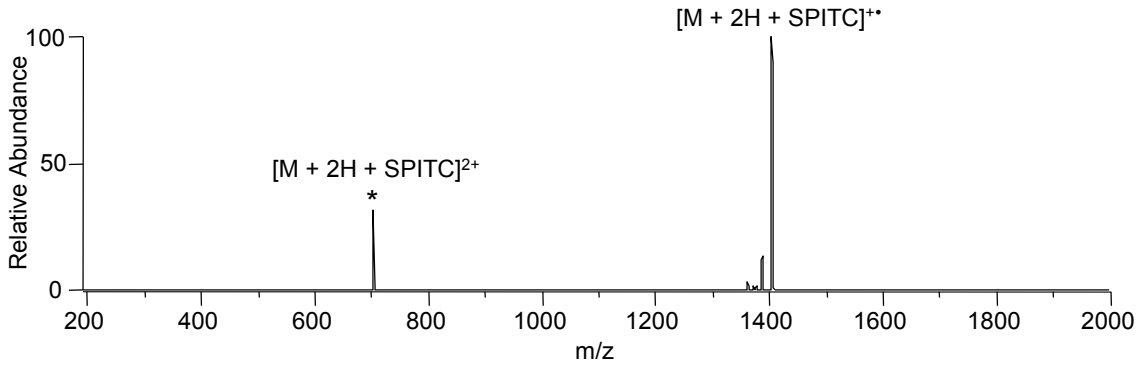

Figure 4.3 Electron transfer dissociation (no supplemental collisional activation) of doubly-charged SPITC-modified lysine-bradykinin. The precursor ion is denoted by *.

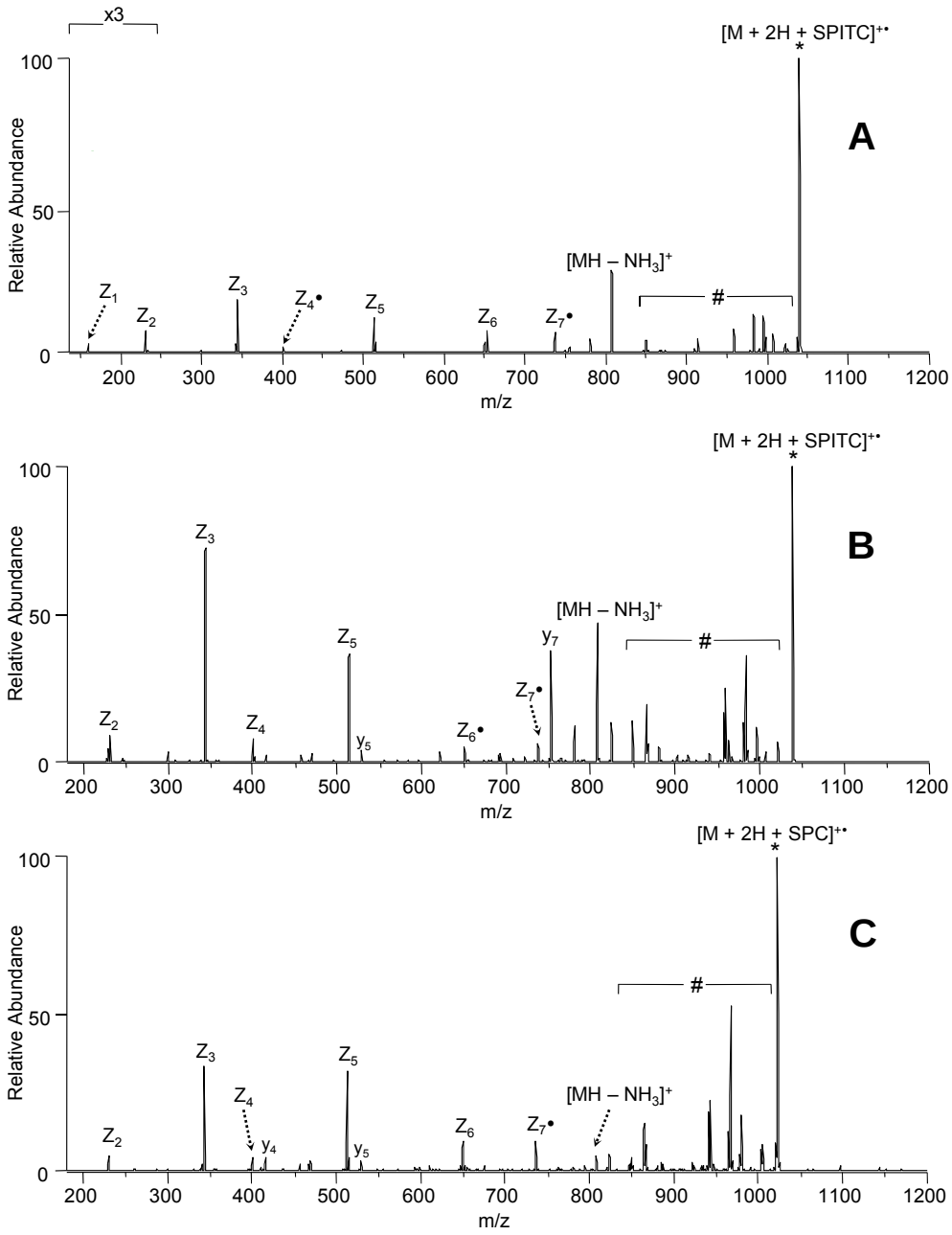

Figure 4.4 ETcaD mass spectra of the doubly charged peptide ASHLGLAR (a) derivatized with SPITC, 24 mV subsequent collision voltage, q = 0.12; (b) derivatized with SPITC, 42 mV subsequent collision voltage, q = 0.16; (c) derivatized with SPC, 42 mV subsequent collision voltage, q = 0.16 . Product ions due to partial decomposition of the SPITC or SPC moiety are denoted by #. The unmodified peptide is denoted by M, and the charge-reduced radical precursor ion is denoted by *.

ETD of SPITC- or SPC-derivatized peptides with more than two charges resulted in significant abundances of both *c*- and *z*-ions and increased sequence coverage than obtained upon ETD of the doubly charged SPITC- or SPC-derivatized peptides, which is analogous to the ETD behavior of unmodified peptides (i.e. more efficient formation of both *c*- and *z*- ions). However, significant abundances of non-dissociated charge-reduced peptides were also observed. Because the ETD spectra for the SPITC- and SPC-modified peptides in the higher charge states display both the N-terminal and C-terminal sequence ions, the resulting spectra contain redundant ions that are less amenable to *de novo* sequencing.

4.4.2 Formation of z_{n-1} versus y_{n-1} Product Ions

It has previously been shown that y_{n-1} is the most abundant product formed upon CID of SPITC-modified peptides, both singly and doubly charged.[14, 15] This phenomenon is illustrated in **Figure 4.2A** in which both the singly and doubly charged y_7 ion dominate the CID mass spectrum obtained for doubly protonated SPITC-ASHLGLAR. This process is promoted by nucleophilic attack of the sulfur atom of the SPITC moiety on the carbonyl oxygen atom of the adjacent amino acid, resulting in an Edman-type degradation.[14, 32] Derivatization of peptides with SPC, which is identical to SPITC except for substitution of a oxygen atom for the sulfur, alleviates the domination of the y_{n-1} ion due to the reduced nucleophilicity of the oxygen.[16] This leads to a more even distribution of *y* ions that is preferable for *de novo* sequencing (**Figure 4.2B**); however, both y and b series were produced upon CID of the doubly protonated, SPC-derivatized peptide which led to numerous redundant ions and a cluttered spectrum. These CID patterns are characteristic of other doubly charged, SPITC- and SPC-derivatized peptides: dominant formation of the y_{n-1} ion for the former and more importantly cluttered spectra for both;

thus neither was ideal for sequencing peptides in mixtures using an LC-ESI-MS^n approach.

In contrast, ETcaD of the SPITC-modified peptides using low supplemental collision voltages yielded a broad series of z ions and did not result in enhanced formation of the y_{n-1} (or corresponding z_{n-1}) product ions, a result that was surprising given the lability of the peptide bond that normally is cleaved to form the y_{n-1} ion upon CID (see **Figure 4.4A** with comparison to **Figure 4.2A**). However, ETcaD using a higher collision voltage produced a more abundant y_{n-1} ion, in addition to the series of *z* ions (**Figure 4.4B**). The y_{n-1} ion is easily identifiable because it is 16 Da greater in mass than the z_{n-1} ion, but its presence can largely be eliminated by using lower energy ETcaD conditions as demonstrated in **Figure 4.4A**. To optimize the formation of the preferred *z*-type ions over the low energy y_{n-1} product, the collision voltage and q-value were varied systematically during the CID activation period. **Figure 4.5** illustrates the resulting trend for formation of the z_{n-1} versus y_{n-1} ions for doubly protonated SPITC-modified ASHLGLAR. As seen in **Figure 4.5**, the z_{n-1} dominates at lower energy ETcaD conditions and the y_{n-1} ion dominates at higher energy ETcaD conditions (i.e. higher q-value and/or greater CID voltage). These experiments parallel Coon's initial ETcaD work where more *b* and *y* ions (instead of *c* and *z*) were seen at higher q-values and greater collision energies.[47] The absence of any *y* product ions using lower energy ETcaD conditions suggests that the radically-driven dissociation pathway that leads to *z*-type ions could be a lower energy process compared to the process that leads to the formation of the y_{n-1} ion.

Another way to suppress the formation of the y_{n-1} ion is to use SPC as the derivatization reagent. **Figure 4.4C** shows the ETcaD mass spectrum of doubly charged SPC-modified ASHLGLAR, and no y_{n-1} ion was formed using any ETcaD conditions;

however some low abundance *y* ions (e.g. y_4 and y_5) were detected at high q-values and CID energies. Because the SPC derivatization procedure involves the use of a less desirable 50% pyridine solution and reaction yields were often low, SPITC was used as the N-terminal derivatizing reagent for the remainder of the study. The SPITC reaction procedure consistently produced the highest yields of modified peptides. For example, guanidination and SPITC derivatization of the tryptic digest of myoglobin (described in more detail in subsequent sections) resulted in 100% conversion of unmodified peptides to modified species.

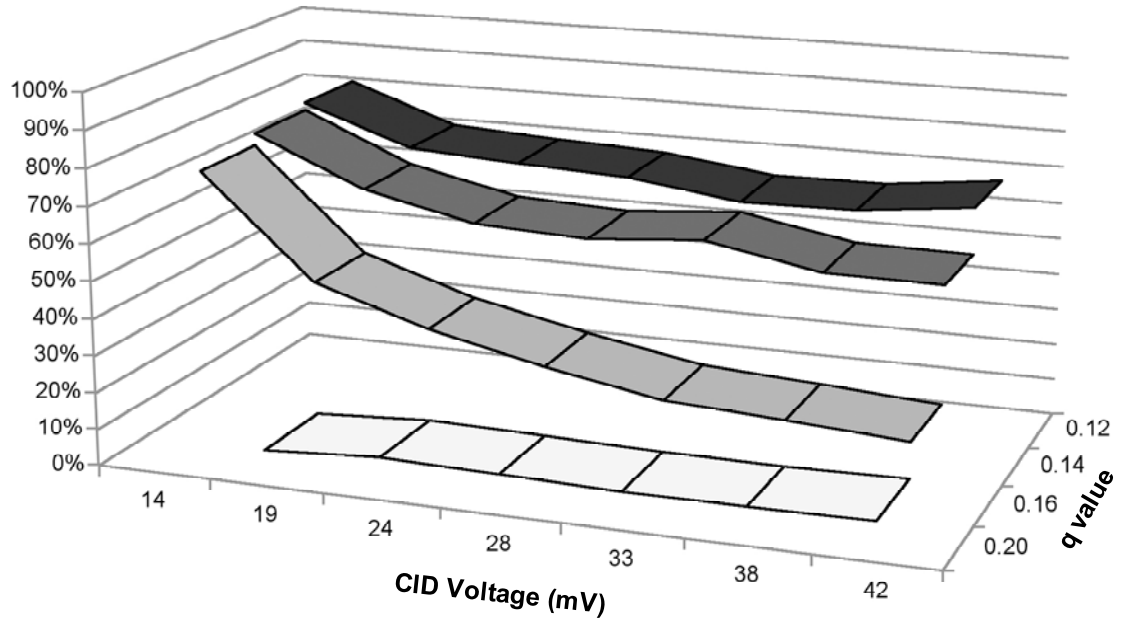

Figure 4.5 Percent abundance of z_{n-1} versus y_{n-1} (i.e. $z_{n-1}/(z_{n-1}+y_{n-1}) \times 100$) in ETcaD spectra as a function of CID voltage and q-value for doubly charged SPITC-ASHLGLAR.

4.4.3 LC-MS^n Analysis of SPITC-Derivatized Tryptic Peptides

To evaluate the analytical utility of the ETcaD method for protein characterization, myoglobin was selected as a model protein for tryptic digestion and SPITC derivatization. Under LC-MS conditions, most of the resulting SPITC-modified tryptic peptides were observed in the 2+ charge state upon ESI. Thus, for data-dependent LC-MS experiments, the doubly charged peptides were typically the ones subjected to ion activation. ETcaD, CID, and ETD spectra were collected in a data-dependent manner for the tryptic peptides in several charge states for separate LC runs of SPITC-derivatized and non-derivatized peptides. The peak areas of the product ions for all C-terminal ions and all N-terminal ions were summed separately and converted to percentages. The results are displayed in **Figure 4.6** to allow ready comparison of the preference for formation of C-terminal (*z* ions) versus N-terminal product ions and other redundant ions such as ones in multiple charge states and redundant C-terminal species. The extreme C-terminal product ion selectivity for the ETcaD spectra obtained for the doubly charged SPITC-peptides is readily apparent, as demonstrated by the percentages of C-terminal ions ranging from 95.3% to 100.0% with an average of 97.1% in **Figure 4.6**. In comparison, CID of the doubly charged SPITC-modified peptides produced more complicated spectra containing both N-terminal and C-terminal ions. Three of the peptides (LFTGHPETLEK, FDKFK, and HLKTEAEMK) appeared to show high C-terminal ion selectivity upon CID, but this was due to production of very dominant y_{n-1} ions and not due to production of a comprehensive series of *y* ions as would be needed for effective *de novo* sequencing. Due to the limited *m/z* range of the ion trap (up to 2000 *m/z*), larger tryptic peptides such as YLEFISDAIIHVLHSK were unable to be analyzed by the ETcaD method after N-terminal derivatization. Although ETD alone and ETD with supplemental CID of unmodified peptides has been shown to increase the level of

sequence coverage and also to promote retention of labile modifications in the product ions,[36, 38, 47] these electron transfer based methods yield spectra that are often much more complicated than traditional CID spectra because highly abundant *c*-, *z*-, and *a*-type ions are present as well as low abundance *y* ions. Thus, the SPITC derivatization procedure in conjunction with ETcaD, not ETD alone, is a key factor that leads to the preferential formation of C-terminal ions. Both ETcaD and CID can be set up as successive scan events during a single data-dependent LC-MS run; therefore, these methods can be used in a complementary rather than competitive mode to increase the overall sequence coverage. In the present study, the average sequence coverage obtained was 86% for all N-terminally derivatized peptides analyzed by ETcaD (shown in **Figure 4.6**) with two major factors that limited the sequence coverage: (1) the low mass cutoff (LMCO) of the ion trap instrument prohibited detection of the lowest *m/z* product ions, and (2) miscleavages from proline residues inherent to ETD. The high sequence coverage obtained was comparable to previous reports of 89% coverage of unmodified peptides by ETcaD.[47] Also, the increased spectral simplification afforded by ETcaD of the SPITC-modified peptides could better unveil miscleavages stemming from the presence of proline residues based on characteristics gaps in the observed series of *z* ions. In summary, the combination of ETcaD with SPITC-derivatization affords the best outcome of high sequence coverage and most simplified spectra containing nearly exclusively *z* ions.

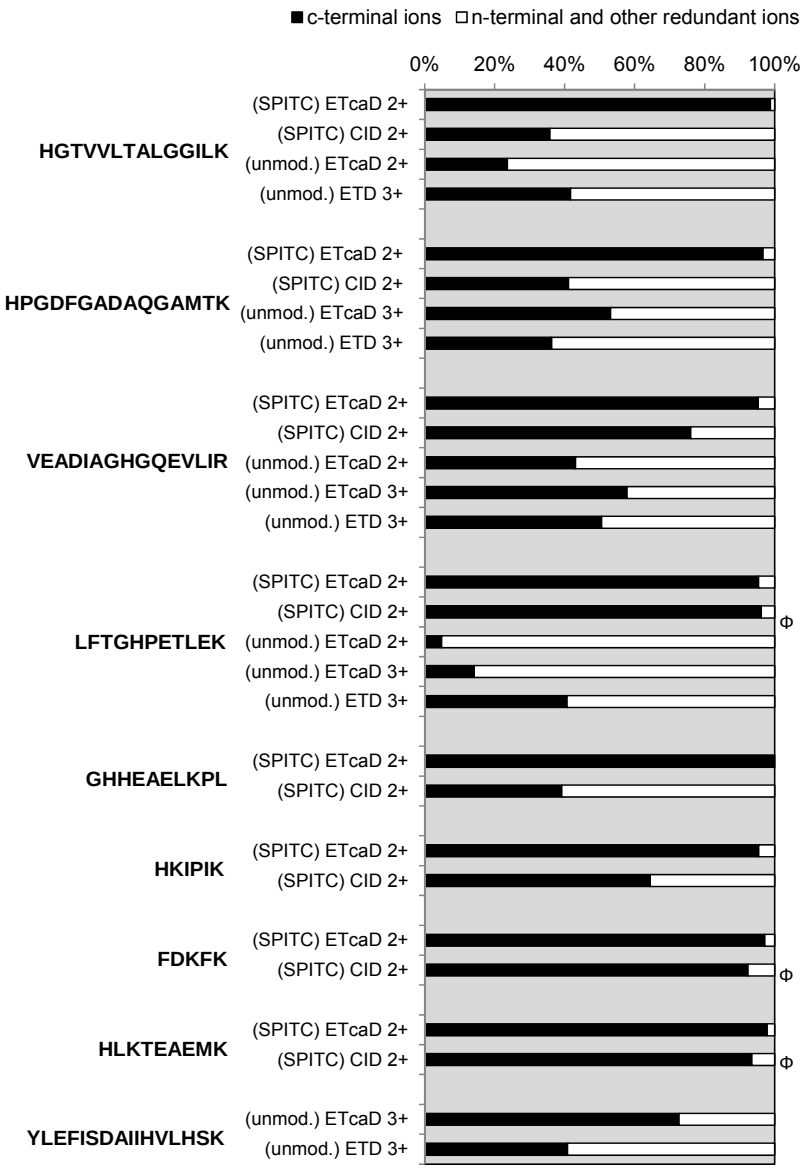

Figure 4.6 Comparison of ETcaD, CID and ETD results for multiply charged tryptic peptides from myoglobin obtained by LC-MS^n. SPITC-derivitized peptides are denoted by (SPITC), and unmodified peptides by (unmod.). The peak areas of the product ions for all C-terminal ions and all N-terminal ions were summed separately and converted to percentages. All lysines are guanidinated for SPITC-modified peptides. Φ denotes peptides with high apparent C-terminal ion selectivity upon CID, but which was actually due to production of very dominant y_{n-1} ions and not due to production of a comprehensive series of *y* ions as would be needed for effective *de novo* sequencing.

4.4.4 Phosphorylated Peptides

One of the biggest advantages of using electron-based dissociation techniques such as ETD is that labile modifications of amino acids generally remain intact, a factor that greatly aids in identifying PTMs.[38-45] However, loss of phosphoric acid is still a substantial dissociation pathway for phosphopeptides examined by ETcaD in a linear ion trap.[47] To circumvent this shortcoming, we evaluated the use of ETcaD for analysis of the N-terminal sulfonated phosphorylated peptide KRpTIRR. The ETcaD spectra for the SPITC-derivatized peptides (i.e. ET of the 2+ and 3+ charge states, then collisional activation of the resulting singly charged radicals; see **Figures 4.7A** and **4.7B**, respectively) displayed an enhanced series of *z* ions with minimal loss of phosphoric acid. Conversely, ETcaD of doubly charged unmodified KRpTIRR (**Figure 4.7C**), CID of singly charged SPITC-modified KRpTIRR (**Figure 4.7D**), and CID of doubly charged SPITC-modified KRpTIRR (**Figure 4.7E**) all showed abundant losses of phosphoric acid and lower sequence coverage.[51] A possible explanation for the minimized loss of phosphoric acid for SPITC-derivatized peptides in comparison to unmodified peptides upon ETcaD is that the loss of SPITC or SPITC + ammonia moieties are even lower energy pathways in comparison to the loss of phosphoric acid, rendering the latter pathway less competitive for the SPITC-derivatized peptides.

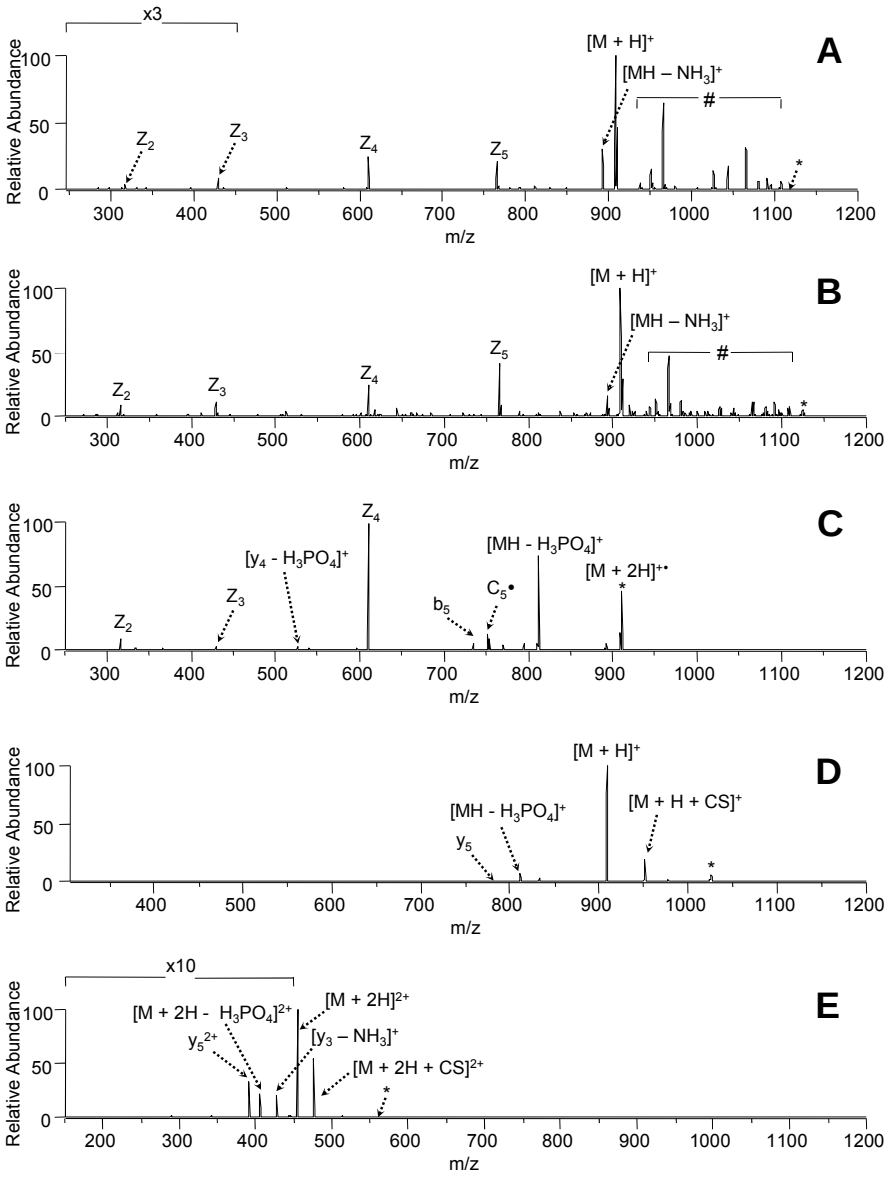

Figure 4.7 ETcaD of the phosphorylated peptide KRpTIRR (a) derivatized with SPITC (doubly charged), 25 mV subsequent collision voltage, q = 0.20; (b) derivatized with SPITC (triply charged)' 30 mV subsequent collision voltage, q = 0.20; (c) unmodified (doubly charged), 21 mV subsequent collision voltage, q = 0.20; and CID of KRpTIRR (d) derivatized with SPITC (singly-charged), 45 mV collision voltage, q = 0.25; (e) derivatized with SPITC (doubly charged), 14 mV collision voltage, q = 0.25. Product ions due to cleavage of a portion of the SPITC moiety are denoted by #, and precursor ions for CID spectra are denoted by *. The unmodified peptide is denoted by M.

The peptide TRDIYETDYpYRK was also N-terminally derivatized with SPITC and subjected to ETcaD analysis. Like the peptide KRpTIRR, SPITC-TRDIYETDYpYRK showed very high sequence coverage upon ETcaD with little neutral loss of phosphoric acid, which is illustrated for the triply charged species in **Figure 4.8**. These results illustrate that the combination of N-terminal derivatization and ETcaD can simultaneously simplify spectra and aid identification of PTMs.

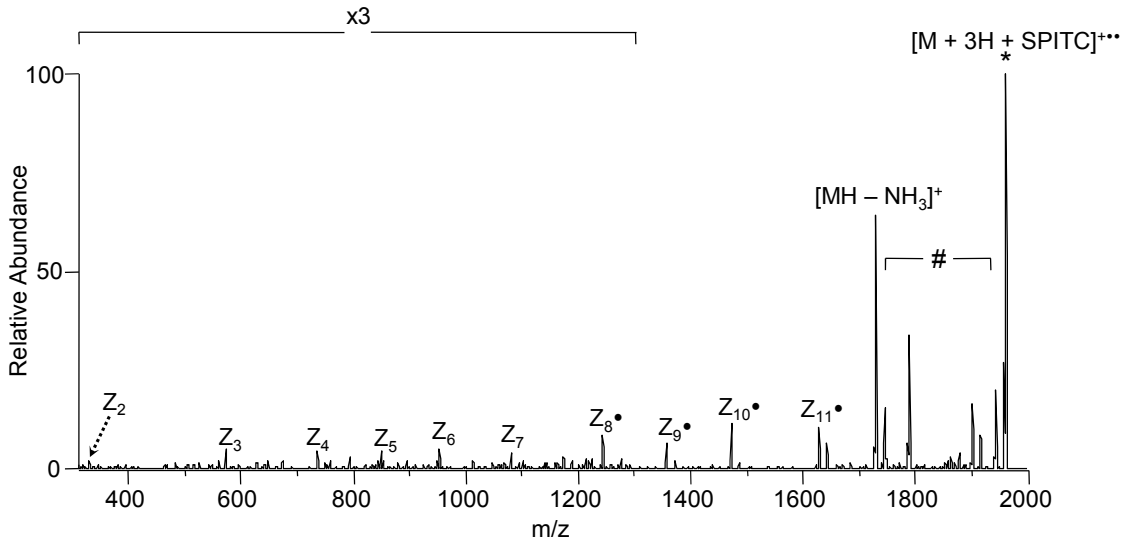

Figure 4.8 ETcaD of the phosphorylated peptide TRDIYETDYpYRK derivatized with SPITC (triply charged), 67 mV subsequent collision voltage, q = 0.14. Product ions due to cleavage of a portion of the SPTIC moiety are denoted by #, the unmodified peptide by M, and the precursor ion for subsequent CID by *.

4.4.5 Benefits of Enhanced *z*- Ions over Enhanced *y*- Ions

Although, ETcaD of multiply charged N-terminally sulfonated peptides and CID of singly-charged N-terminally sulfonated peptides are complementary methods for data-dependent LC-MS analyses, there are significant advantages of the former method. One

obvious advantage is that the more highly charged, larger peptides provide greater protein sequence coverage, and their fragmentation patterns can only be successfully simplified by the ETcaD technique. Also as previously described, the *z* ions produced during the ETcaD technique retain labile PTMs. In contrast, CID of the even-electron, singly-charged SPITC-derivatized peptides produced *y* ions which did not retain PTMs and offered lower sequence coverage for phosphopeptides. Furthermore, a highly basic residue (e.g. arginine or homoarginine) is needed at the C-terminus to enhance the *y* ions for CID analysis of SPITC-modified peptides.[14] Without these highly basic residues, the proton is not sequestered on the C-terminus (i.e. it is more mobile), resulting in a more diverse array of both C-terminus and N-terminus ions and lower sequence coverage. However, a highly basic arginine or homoarginine is not needed at the C-terminus when using ETcaD due to the different radical dissociation mechanism inherent to electron-based methods, yet the resulting spectra still offer high sequence coverage and yield a single dominant series of product ions. This advantage has already been seen for the peptide GHHEAELKPL in **Figure 4.6**. Therefore, the combination of ETcaD and N-terminal derivatization could potentially be used for spectral simplification of a variety of peptide mixtures created from proteins using enzymes that do not necessarily cleave after lysines and arginines.

4.4.6 Formation of Even Electron versus Radical *z* Ions

Since both even electron and radical *z* product ions are formed by ETcaD, it was relevant to assess the formation preference of these two types of *z* ions. These two types can generally be distinguished from each other as both the even electron and radical ion peaks are usually present at each *z* ion position and differ by one mass unit. However, the ability to predict the preference for formation of the radical or even electron *z* ions would

enhance the success of *de novo* sequencing algorithms since some amino acids differ in mass by only one amu. To gain insight into the *z* ion distributions, area counts of all even electron and radical *z* ions were tabulated for each *z* ion (z_1, z_2, z_3…z_{14}) from 15 different peptide precursor ions corresponding to a total of 23 different spectra, which included different charge states and two different CID q-values and collision voltages. Both SPITC-modified model peptides and SPITC-modified peptides from tryptic digests of myoglobin and cytochrome c were included in this facet of the study. The percentages of even electron *z* ions relative to the total amount of even electron and radical *z* ions are summarized in **Figure 4.9** for each z_n product. As illustrated in **Figure 4.9**, there is a notable increase in the formation of even electron *z* ions as the length of the *z* ion decreases. The identity of specific amino acids adjacent to the cleaved amide bond had little or no impact on the extent of even electron *z* ion formation. However, further investigation with large data sets may be warranted. There was also no trend in the formation of one type of *z* ion when using lower or higher energy ETcaD conditions.

The proposed mechanism for the formation of even electron *z* species from *z* radical species has been described previously, and it involves the formation of a hydrogen bond between the *z* radical and a *c* even electron species (either in an intramolecular or intermolecular process).[52] The *z* radical strips a hydrogen atom from the *c* ion, forming an even electron *z* ion and a radical *c* product.[52] This pathway is consistent with the trend seen in **Figure 4.9**: the longer *z* ions may have greater opportunity to fold and form stabilizing intramolecular hydrogen bonds, thus suppressing hydrogen bond formation with *c* ions. For the case of SPITC-derivatized peptides, the *c* products expected upon separation of the *z* and *c* species are neutralized instead of remaining positively charged.

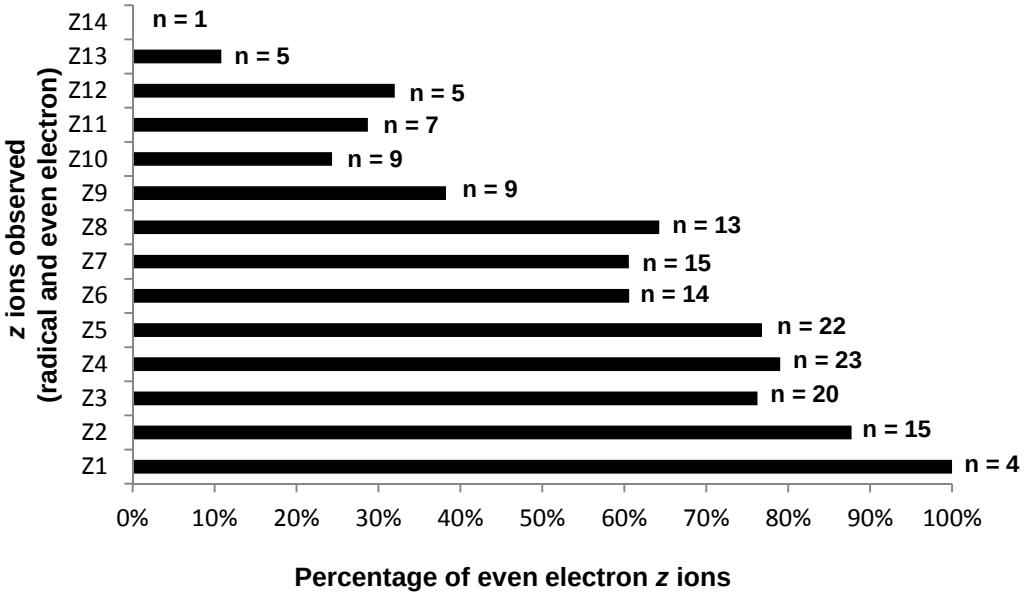

Figure 4.9 Percentage of even electron versus radical *z* ions as a function of z ion length for 15 different peptide precursor ions and a total of 23 different spectra. n = the total number of peptides that had observable *z* ions at each specified sequence position.

4.5 Conclusions

ETcaD of multiply charged, N-terminally sulfonated peptides results in significant spectral simplification via almost exclusive formation of *z*-type ions. This method offers advantages not only compared to CID of multiply charged N-terminally derivatized species, but also to electron transfer based methods (i.e. ETD, ETcaD) of unmodified peptides. The enhanced series of *z* ions was shown to retain labile phosphoric acid moieties and also allow efficient sequencing of a broad range of peptides, including proline-rich ones and non-tryptic species. High sequence coverage was attained for all peptides with the largest limiting factor being the low mass cutoff (LMCO) inherent to the ion trap, which sometimes limited the detection of the lowest *m/z* ions. By using a

photodissociation method, such as IRMPD instead of CID during the subsequent activation step, it is anticipated that the LMCO problem could be alleviated.

4.6 References

(1) Service, R. F. *Science* **2008**, *321*, 1758-1761.
(2) Eng, J. K.; McCormack, A. L.; Yates, J. R., III *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 976-989.
(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.
(4) Taylor, J. A.; Johnson, R. S. *Analytical Chemistry* **2001**, *73*, 2594-2604.
(5) Johnson, R. S.; Taylor, J. A. *Molecular Biotechnology* **2002**, *22*, 301-315.
(6) Taylor, J. A.; Johnson, R. S. *Rapid Communications in Mass Spectrometry* **1997**, *11*, 1067-1075.
(7) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *Journal of Computational Biology* **1999**, *6*, 327-342.
(8) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 2337-2342.
(9) Mann, M.; Wilm, M. *Analytical Chemistry* **1994**, *66*, 4390-4399.
(10) Sergeant, K.; Samyn, B.; Debyser, G.; Van Beeumen, J. *Proteomics* **2005**, *5*, 2369-2380.
(11) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 6855-6862.
(12) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 7131-7136.
(13) Marekov, L. N.; Steinert, P. M. *Journal of Mass Spectrometry* **2003**, *38*, 373-377.
(14) Lee, Y. H.; Kim, M.-S.; Choie, W.-S.; Min, H.-K.; Lee, S.-W. *Proteomics* **2004**, *4*, 1684-1694.
(15) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 3019-3027.
(16) Shin, J.-W.; Lee, Y. H.; Hwang, S.; Lee, S.-W. *Journal of Mass Spectrometry* **2007**, *42*, 380-388.
(17) Samyn, B.; Debyser, G.; Sergeant, K.; Devreese, B.; Van Beeumen, J. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 1838-1852.
(18) Wang, D.; Kalume, D.; Pickart, C.; Pandey, A.; Cotter, R. J. *Analytical Chemistry* **2006**, *78*, 3681-3687.
(19) Wang, D.; Cotter, R. J. *Analytical Chemistry* **2005**, *77*, 1458-1466.
(20) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 2348-2356.
(21) Wang, D.; Kalb Suzanne, R.; Cotter Robert, J. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 96-102.
(22) Wang, D.; Xu, W.; McGrath, S. C.; Patterson, C.; Neckers, L.; Cotter, R. J. *Journal of Proteome Research* **2005**, *4*, 1554-1560.
(23) Pashkova, A.; Chen, H.-S.; Rejtar, T.; Zang, X.; Giese, R.; Andreev, V.; Moskovets, E.; Karger, B. L. *Analytical Chemistry* **2005**, *77*, 2085-2096.

(24) Oehlers, L. P.; Perez, A. N.; Walter, R. B. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 752-758.
(25) Guillaume, E.; Panchaud, A.; Affolter, M.; Desvergnes, V.; Kussmann, M. *Proteomics* **2006**, *6*, 2338-2349.
(26) Bauer, M. D.; Sun, Y.; Keough, T.; Lacey, M. P. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 924-929.
(27) Bao, J.; Ai, H.; Fu, H.; Jiang, Y.; Zhao, Y.; Huang, C. *Journal of Mass Spectrometry* **2005**, *40*, 772-776.
(28) Wang, F.; Fu, H.; Jiang, Y.; Zhao, Y. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 995-999.
(29) Beardsley, R. L.; Reilly, J. P. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 158-167.
(30) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. *Analytical Chemistry* **2005**, *77*, 6300-6309.
(31) Brancia, F. L.; Butt, A.; Beynon, R. J.; Hubbard, S. J.; Gaskell, S. J.; Oliver, S. G. *Electrophoresis* **2001**, *22*, 552-559.
(32) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. *Journal of Mass Spectrometry* **1997**, *32*, 225-231.
(33) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.-a.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. *Analytical Chemistry* **2006**, *78*, 7861-7869.
(34) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2007**, *79*, 7883-7892.
(35) Vasicek, L. A.; Wison, J. J.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 377-384.
(36) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 9243-9252.
(37) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.
(38) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(39) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Analytical Chemistry* **1999**, *71*, 4431-4436.
(40) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 1793-1800.
(41) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Analytical Chemistry* **2001**, *73*, 4530-4536.
(42) Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 7450-7456.
(43) Srikanth, R.; Wilson, J.; Bridgewater, J. D.; Numbers, J. R.; Lim, J.; Olbris, M. R.; Kettani, A.; Vachet, R. W. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1499-1506.

(44) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. *Molecular and Cellular Proteomics* **2008**, *7*, 1838-1849.
(45) Sweet, S. M. M.; Mardakheh, F. K.; Ryan, K. J. P.; Langton, A. J.; Heath, J. K.; Cooper, H. J. *Analytical Chemistry (Washington, DC, United States)* **2008**, *80*, 6650-6657.
(46) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. *Analytical Chemistry* **2008**, *80*, 4825-4835.
(47) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Analytical Chemistry* **2007**, *79*, 477-485.
(48) Han, H.; Xia, Y.; McLuckey, S. A. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 1567-1573.
(49) Xia, Y.; Han, H.; McLuckey, S. A. *Analytical Chemistry* **2008**, *80*, 1111-1117.
(50) Wang, D.; Kalb, S. R.; Cotter, R. J. *Rapid Communications in Mass Spectrometry* **2003**, *18*, 96-102.
(51) guanidination, K.
(52) O'Connor, P. B.; Lin, C.; Cournoyer, J. J.; Pittman, J. L.; Belyayev, M.; Budnik, B. A. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 576-585.

Chapter 5

Top-Down Protein Fragmentation by Infrared Multiphoton Dissociation in a Dual Pressure Linear Ion Trap

5.1 Overview

Infrared multiphoton dissociation (IRMPD) was implemented in a novel dual pressure linear ion trap for rapid top-down proteomics. The high pressure cell provided improved trapping and isolation efficiencies while the isotopic profiles of 10+ charged ions could be resolved by mass analysis in the low pressure cell that enabled effective top down protein identification. Striking differences between IRMPD in the low pressure cell and CID in the high pressure cell were observed for proteins ranging from 8.6 kDa to 29 kDa. Due to secondary dissociation, IRMPD yielded product ions in significantly lower charge states as compared to CID, thus facilitating more accurate mass identification and streamlining product ion assignment. This outcome was especially useful for database searching of larger proteins (~29 kDa) as IRMPD substantially improved protein identification and scoring confidence. Also, IRMPD showed an increased selectivity towards backbone cleavages N-terminal to proline and C-terminal to acidic residues (especially for the lowest charge states), which could be useful for a priori spectral predictions and enhanced database searching for protein identification.

5.2 Introduction

The most common proteomics technique for protein analysis uses a bottom-up mass spectrometry strategy.[1, 2] Here, proteins are extracted from cells, enzymatically

digested, separated, ionized, and analyzed by MS and MS/MS. Database search algorithms are then employed to correlate MS/MS spectra with sequence and ultimately the parent protein. The top-down strategy, an alternative to the bottom-up approach, relies on fragmentation of intact protein cations in the gas phase (i.e., no enzymatic digestion). Although more challenging, top-down methods offer certain benefits over the bottom-up approaches including the elimination of enzymatic digestion procedures, access to contextual post translational modification (PTM) information, and direct knowledge about the molecular weights of the intact proteins.[3-5] To date, top-down proteomic applications have mainly been undertaken on FT-ICR instruments,[3] which can resolve highly charged product ions created upon collision induced dissociation (CID),[6-8] electron capture dissociation (ECD),[9-11] infrared multiphoton dissociation (IRMPD),[12] or ultraviolet photodissociation (UVPD)[13] of the intact proteins. Top-down applications utilizing various hybrid instrument platforms such as the linear ion trap (LIT)/FT-ICR,[4] LIT/time-of-flight,[5, 14] and LIT/orbitrap[15] that allow physical separation of the ion activation and mass analysis steps have gained momentum.

Recently, some inroads in top-down proteomics have been achieved using cheaper, faster, and more robust quadrupole ion trap instruments.[12-18] For instance, McLuckey and coworkers have shown successful protein identification by dissociation of whole proteins by CID followed by gas-phase ion/ion reactions undertaken to convert multiply charged product ions into singly charged ions for easier product ion assignment.[16-19] Other groups have reported top-down protein analysis in quadrupole ion traps by combining electron transfer dissociation (ETD) with other ion/ion reactions[20] or by using extended ETD periods.[21] Although these ion trap techniques have shown substantial promise in the identification of proteins, they generally require relatively long activation/reaction times (up to 300 ms)[21] or require greater abundances of precursor ions

than conventional CID.[18] Some of the resulting singly charged product ions that are useful for protein identification may have *m/z* values beyond the range of a typical quadrupole ion trap (i.e., greater than *m/z* 2000) due to the subsequent proton transfer reactions.[18]

IRMPD is another useful tandem MS method that allows trapping and detection of a wide *m/z* range because the trapping conditions are decoupled from the activation process,[22] IRMPD promotes secondary dissociation,[12, 23, 24] and IRMPD can be used for confident identification of some types of post-translational modifications (PTMs).[23, 25-28] Due to the substantial competition between IR energization and collisional cooling that occurs at typical operating pressures of quadrupole ion traps (~1 mtorr), IRMPD has had limited implementation in quadrupole ion trap instruments. Our group and others have investigated methods to improve fragmentation efficiencies in these instruments by attaching IR absorbing chromophores,[29-32] focusing the laser source,[33] supercharging the ions of interest prior to activation,[24] thermal assistance,[34] and more recently utilizing a dual pressure linear ion trap.[35] However, efficient top-down proteomics using IRMPD in quadrupole ion traps until now has remained largely unexplored. Due to the inherent lower pressures of FT-ICR and linear ion trap/time-of-flight mass analyzers, both McLafferty and Siu have used these instruments in conjunction with IRMPD for successful top-down analysis.[12, 36] In both of these prior studies, the dominant IRMPD pathways involved cleavages of the protein backbone on the C-terminal side of glutamic and aspartic acid residues as well as on the N-terminal side of proline residues.[12, 36]

Another formidable challenge of top-down proteomics stems from the lack of knowledge about the gas-phase fragmentation behavior of whole protein cations.[37] Improvements to a priori spectral predictions and incorporation into searching algorithms could result in more confident protein identification. Therefore, strategies have been

developed by Julian and coworkers to direct backbone fragmentation of intact proteins at specific residues,[38] but this approach necessitates chemical derivatization. Others have reported an increased selectivity of backbone cleavages N-terminal to proline and C-terminal to acidic residues (glutamic acid and aspartic acid) after CID of whole proteins,[39-43] but this preferential cleavage has been difficult to predict a priori and depends heavily on precursor charge state. For example, it has been shown that at low precursor charge states, cleavages C-terminal to aspartic acid, and to a lesser extend C-terminal to glutamic acid, are prevalent.[40] Nonspecific cleavages have shown to dominate at intermediate charge states, and cleavages N-terminal to proline are abundant at higher charge states.[40]

In the present study, IRMPD is implemented in a dual pressure linear ion trap for top-down protein analysis. Fragmentation patterns (i.e., charge state distributions and backbone cleavage selectivity) of proteins after IRMPD in the low pressure cell are compared to CID results in the high pressure cell. Further comparisons between the two dissociation methods are made from *in silico* database searches.

5.3 Experimental

5.3.1 Materials

Myoglobin and cytochrome c from equine (*Equus caballus*) heart, bovine (*Bos Taurus*) carbonic anhydrase, ubiquitin, and α-SI-casein, and all other reagents were purchased from Sigma-Aldrich (St. Louis, MO). Proteins were analyzed without further purification. Working solutions of 10 μM proteins in 49.5:49.5:1.0 water/methanol/acetic acid (v/v/v) were prepared and directly infused at a flow rate of 3 μL/min into the mass spectrometer; ions were formed through electrospray ionization.

5.3.2 Mass Spectrometry and Infrared Multiphoton Dissociation

Proteins were analyzed using a modified LTQ Velos, which is a differentially pumped dual pressure linear ion trap (Thermo Fisher, San Jose, CA).[44] The high pressure cell was operated at nominally 4.7 mTorr, and the low pressure cell at nominally 0.3 mTorr. Protein ions are accumulated in the high pressure cell for more efficient trapping and isolation, and were then transferred to the low pressure cell where m/z analysis was performed in zoom scan mode. The automatic gain control (AGC) was set to 500, and the ESI voltage to 4.5 kV. Sheath and auxiliary nitrogen gas flow rates were 0.12 L/min and 0.3 L/min, respectively. The heated capillary temperature was maintained at 200ºC. IRMPD experiments were performed in both the high and low pressure cells using a Synrad 50 W continuous wave CO_2 laser (model 48-5; Mukilteo, WA) operated at full power (~50 W). Laser radiation at a wavelength of 10.6 μm was introduced into the linear ion trap via a modified back plate containing a ZnSe window.[32] The precursor m/z was placed at a q-value of 0.1 for all IRMPD experiments, and irradiation times ranged from 5.0 – 18.0 ms. CID experiments were performed in the high pressure cell to maximize dissociation efficiencies. Standard parameters (q-value of 0.25 and activation times of 30 ms) were used for all CID experiments. Further IRMPD and MS^3 experiments were performed on a modified, LTQ XL linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). All spectra were generated from 20 – 34 averages of three microscans.

5.3.3 Data Analysis

The charge states of product ions were identified by inspection of the isotopic envelope. Sequence ions were identified when two criteria were met: when appropriate isotopic envelopes could be distinguished and the ion peaks had signal:noise ratios equal

to or greater than three. Origin 7.0 was used to compute peak areas and intensities for further MS/MS data interpretation. Dissociation efficiencies were calculated based on the percentage of the original precursor ions (based on peak areas) which were converted into product ions. A value of 100% signified that the initial area of the isolated precursor ion peak was completely accounted for by the summation of all the areas of the product ion peaks. ProsightPTM 2.0[45-47] was used for database searching of *Bos taurus* proteins (e.g., ubiquitin and carbonic anhydrase). Peaks were picked by Origin 7.0, and charge states were manually assigned. Excel was used to generate zero-charge (deconvoluted) spectra by multiplying the m/z values (from Origin) by assigned charge states and subtracting appropriate protons. The spectra were input into ProsightPTM in its absolute mass mode using average masses, and a precursor mass window of 2000 Da and a fragment tolerance of 1.5 Da were used for all searching. Probability scores (p-score) are based on a Poisson distribution that reflects the probability of matching identified product ions to a given protein sequence; scores equal to or lower than 0.05 are statistically significant.[14, 48] Expectation values equal the probability score multiplied by the number of protein sequences in the *Bos taurus* database (i.e. expectation value = p-score x 2,869,593 protein forms).[49] The PDE score, which is based on the McLuckey score,[50] takes into account product ion intensities and places weighting factors that increase scores when product ions stem from cleavages N-terminal to proline (5x weighting factor), C-terminal to aspartic acid (5x weighting factor), C-terminal to lysine (4x weighting factor), C-terminal to glutamic acid (2x weighting factor), N-terminal to proline and C-terminal to aspartic acid (10x weighting factor), N-terminal to proline and C-terminal to lysine (9x weighting factor), and N-terminal to proline and C-terminal to glutamic acid (7x weighting factor). Higher PDE scores reflect better confidences in protein identification.

5.4 Results and Discussion

In this chapter, proteins ranging from 8.6 kDa to 29 kDa (e.g., ubiquitin, myoglobin, cytochrome C, and carbonic anhydrase) were ionized by ESI and injected into a dual pressure linear ion trap for mass analysis. Ions were accumulated in the high pressure cell for more efficient trapping and isolation, and were then transferred to the low pressure cell for high resolution mass analysis. For MS/MS experiments, proteins could be dissociated in either cell. All product ion spectra were acquired using the zoom scan rate in which ions up to the 10+ charge state could be isotopically resolved with peak widths of ~0.1 u (FWHM) at *m/z* 1000 (**Figure 5.1**). By performing mass analysis in the low pressure cell, resolution is considerably improved over that obtained in traditional high pressure linear ion traps, thus enabling the resolution and ready identification of significantly higher charged product ions.

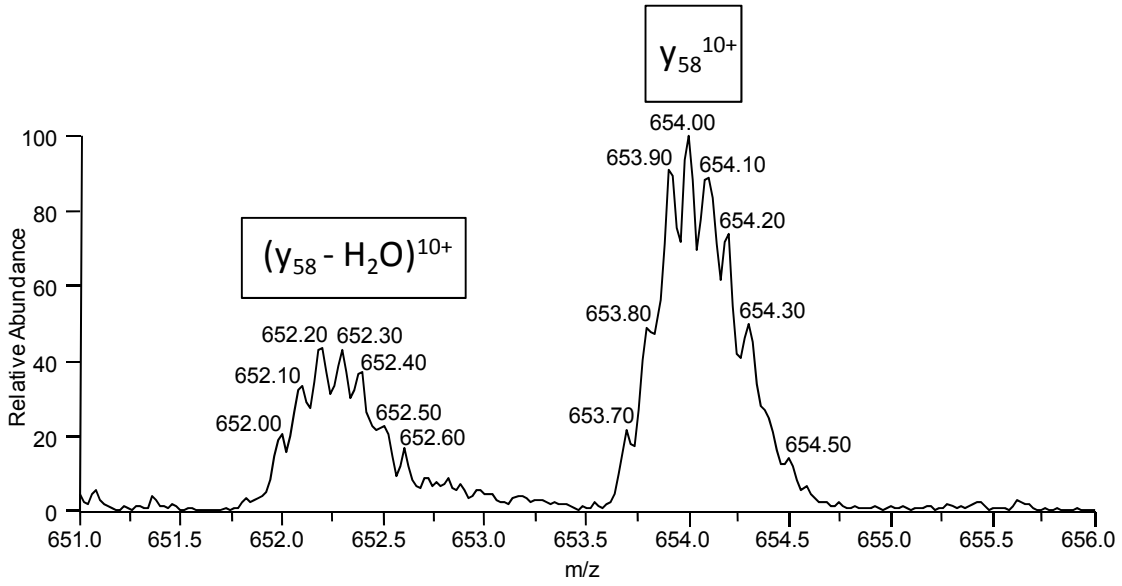

Figure 5.1 Expansion of isotopic profiles of 10+ product ions.

5.4.1 IRMPD Protein Dissociation Efficiency

The dissociation efficiencies for various charge states of ubiquitin (7+ through 12+) produced by IRMPD in the low pressure cell and by CID in the high pressure cell were compared. In general, both activation methods yielded high dissociation efficiencies (i.e., near 80% on average of all charge states analyzed), although the dissociation efficiency was slightly lower for IRMPD compared to CID (78 $\pm$ 6% for IRMPD versus 86 $\pm$ 4% for CID). The higher CID efficiency was attributed to the more efficient energization of ions in the high pressure cell.[44] The improved fragmentation observed for IRMPD using the low pressure cell of the linear trap, particularly for the intact proteins targeted in this study, was a vast improvement over that observed in traditional, single pressure, linear ion traps.

5.4.2 Protein Fragmentation

The efficient dissociation of proteins by IRMPD in the low pressure cell is illustrated in **Figure 5.2A** for the 7+ charge state of ubiquitin (irradiation time of 18 ms). The dominant product ions were those from selective backbone cleavages N-terminal to proline ("P") or C-terminal to acidic residues ("D" for aspartic acid and "E" for glutamic acid). Product ions evolving from backbone cleavage between an adjacent acidic residue (C-terminal) and proline (N-terminal) were denoted by either "E/P" or "D/P." Product ions resulting from all other nonselective cleavages (i.e., ones producing other a, b, and y ions) were denoted by $\otimes$. The corresponding CID mass spectrum of the 7+ charge state of ubiquitin is shown in **Figure 5.2B**. Comparison of the spectra in **Figures 5.2**, a (IRMPD) and b (CID), reveal several differences. For instance, the majority of the product ions formed upon IRMPD were observed in significantly lower charge states than those produced by CID, thus enabling easier interpretation of isotopic profiles and higher mass

accuracy assignments. This phenomenon was observed previously for IRMPD of peptides in a dual pressure linear trap[35] and was attributed to secondary dissociation of the more highly charged product ions. Also, due to secondary dissociation the distributions of product ions observed in the IRMPD and CID mass spectra were considerably different. For example, the most abundant sequence ions (e.g., b and y ions) in the IRMPD mass spectrum included those stemming from backbone cleavages C-terminal to aspartic acid residues (D), whereas for CID the most abundant products arose from backbone cleavages between adjacent glutamic acid (C-terminal) and proline (N-terminal) residues as well as those from nonselective backbone cleavages (other a, b, and y ions). The broader *m/z* trapping range possible during IRMPD also enabled the identification of low mass product ions, ones which could not be detected upon CID due to the low mass cutoff established by the rf voltage used during ion activation (e.g., b_2 and b_3). Some internal ions of low abundance resulting from backbone cleavages N-terminal to proline and C-terminal to acidic residues were also observed upon IRMPD.

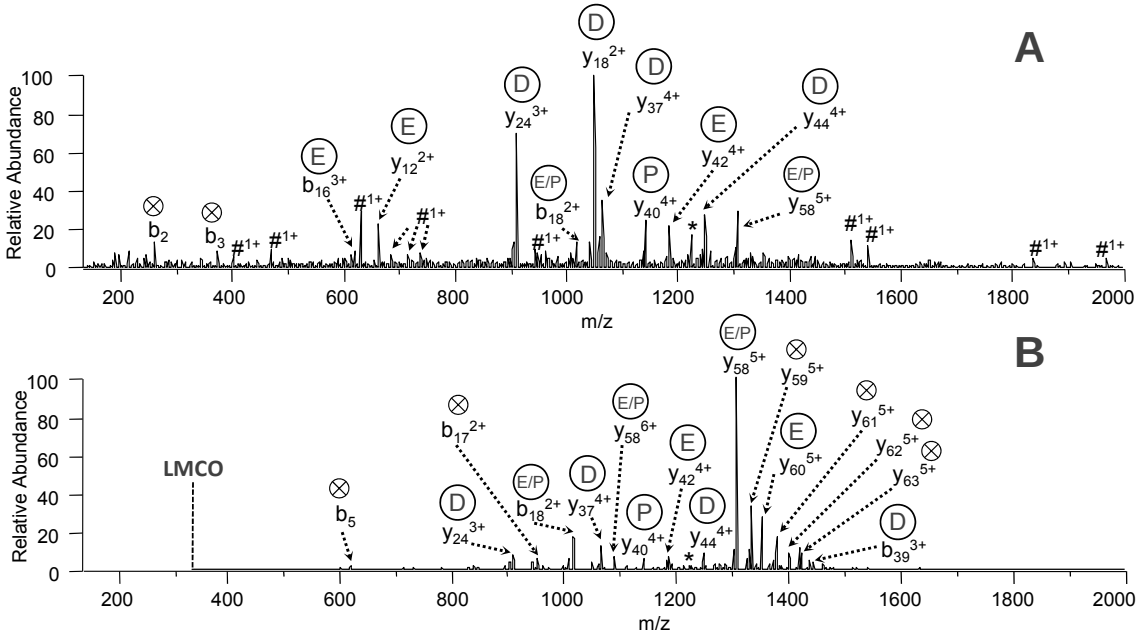

Figure 5.2 Ubiquitin, MW 8,565 Da $[M + 7H]^{7+}$ (*m/z* 1224.5) dissociated by (a) IRMPD in the low pressure cell, t = 18.0 ms and (b) CID in the high pressure cell, 21% NCE. Product ions resulting from selective cleavages N-terminal to proline and C-terminal to acidic residues are denoted by P and D, E (respectively). Internal ions resulting from these selective cleavages are denoted by #. Product ions resulting from unselective cleavages are denoted by ⊗, and precursor ions by *.

Another example is shown in **Figure 5.3** for the analysis of myoglobin in the 14+ charge state by IRMPD (**Figure 5.3A**) and CID (**Figure 5.3B**). Again, lower product ion charge states and higher abundances of ions attributed to selective backbone cleavages N-terminal to proline and C-terminal to acidic residues are observed upon IRMPD (15 ms). In general, the IRMPD spectrum was much less complex and thus easier to interpret. Conversely, CID resulted a notably more cluttered spectrum with several unidentifiable ions in charge states too high to resolve the isotopic distribution (denoted by "?") under the mass analysis resolution conditions achieved using the experimental conditions described in the methods section above. Moreover, the most abundant products and in

fact more than half of the total product ions from CID resulted from nonselective backbone cleavages. CID did, however, generate a greater array of product ions, albeit at the expense of low mass sequence ions, such as the b_2 ion, and low mass internal ions, that serve as useful fingerprint ions when identifying intact proteins.

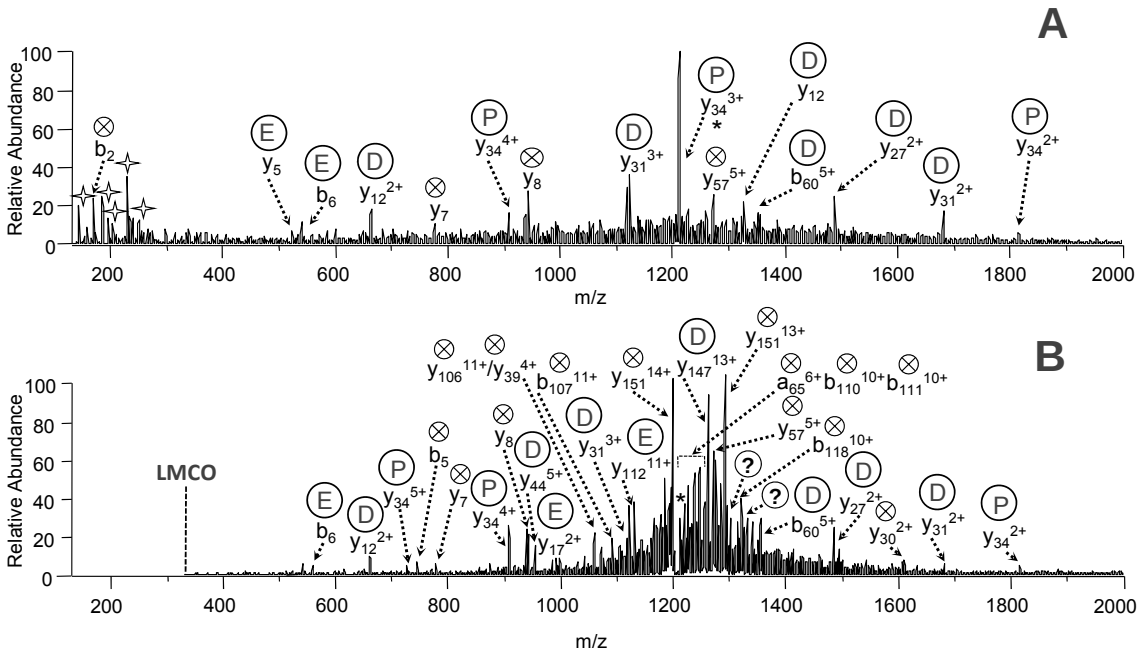

Figure 5.3 Myoglobin, MW 16,951 Da $[M + 14H]^{14+}$ (*m/z* 1211.7) dissociated by (a) IRMPD in the low pressure cell, t = 15.0 ms and (b) CID in the high pressure cell, 22% NCE. Product ions resulting from selective cleavages N-terminal to proline and C-terminal to acidic residues are denoted by P and D, E (respectively). Small internal ions of two amino acids are denoted by ✧. Product ions resulting from unselective cleavages are denoted by ⊗, and precursor ions by *.

Even for higher molecular weight proteins in high precursor charge states, IRMPD yields product ions in low charge states with a high selectivity for backbone cleavages occurring N-terminal to proline and C-terminal to acidic residues. This general

trend is illustrated in **Figure 5.4A**, which shows the IRMPD spectrum of the 20+ charge state of carbonic anhydrase (29 kDa) at a short irradiation time of only 9 ms. Several low mass sequence ions (e.g., y_2, b_2, b_3, and b_4) were again observed for IRMPD that were not detected upon CID (**Figure 5.4B**). Both dissociation methods generated the same number of identified sequence ions; however, CID had many more highly charged ions that were neither resolved nor identified. Also, as seen previously, a greater portion of the product ions were attributed to ones resulting from nonselective backbone cleavages for CID as compared to IRMPD, a general observation described in more quantitative detail later.

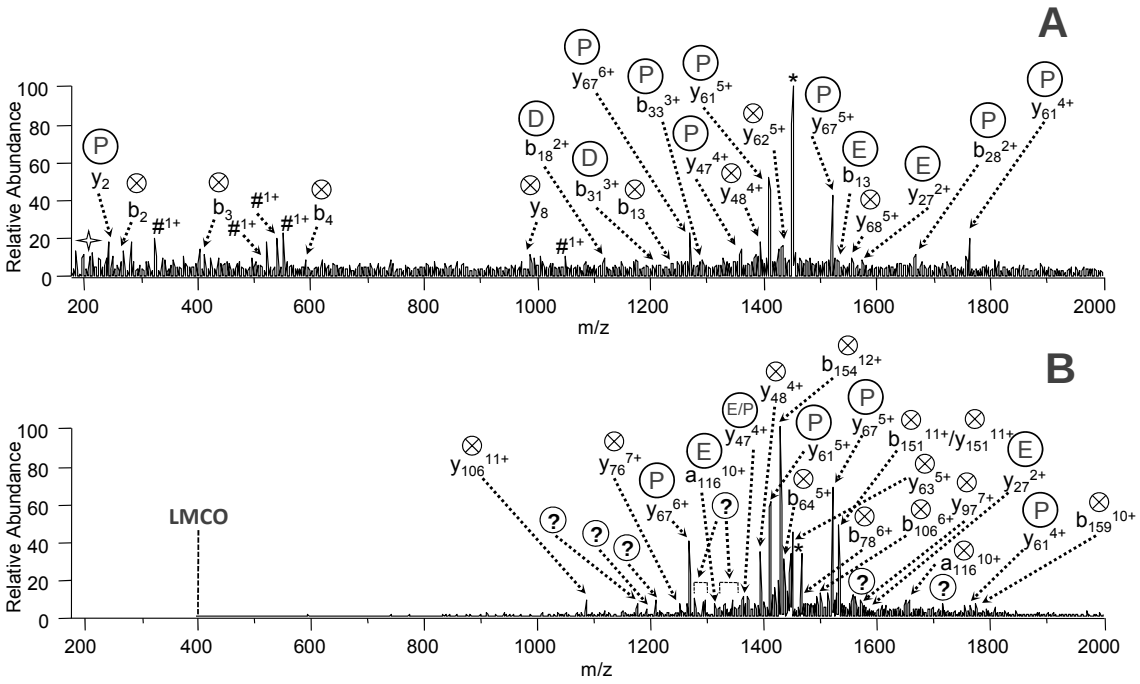

Figure 5.4 Carbonic anhydrase, MW 29,025 Da $[M + 20H]^{20+}$ (*m/z* 1452.4) dissociated by (a) IRMPD in the low pressure cell, t = 9.0 ms and (b) CID in the high pressure cell, 10% NCE. Product ions resulting from selective cleavages N-terminal to proline and C-terminal to acidic residues are denoted by P and D, E (respectively). Internal ions resulting from these selective cleavages are denoted by #, and small internal ions of two amino acids are denoted by ✧. Product ions resulting from unselective cleavages are denoted by ⊗, and precursor ions by *.

Dissociation of the 10+ charge state of cytochrome C by IRMPD (**Figure 5.5A**) and CID (**Figure 5.5B**) resulted in large product ion signals from the detachment of heme from the protein backbone; however, major spectral differences were still observed. Again, lower charged product ions were seen for IRMPD as compared to CID. In fact, for IRMPD at an irradiation time of 11.0 ms (**Figure 5.5A**), 68% of the total abundance of product ions (based on integrated peak areas) resulted from singly charged ions. This

surprisingly low charge state distribution was rather unexpected, but could be attributed to increased IR absorptivity from the covalently attached heme group that promotes even more extensive secondary dissociation relative to unmodified (non-heme) proteins. Also due to secondary dissociation, the anticipated $[M - Heme]^{9+}$ ion was not detected upon IRMPD. This $[M - Heme]^{9+}$ ion was highly abundant in the CID spectrum and proved to be a dead-end (noninformative) product that provided no diagnostic sequence information. CID did generate, however, a greater array of sequence ions as compared to IRMPD, albeit few low mass sequence ions due to the intrinsic narrow m/z range characteristic of the trapping conditions typically employed for CID. Both dissociation methods yielded product ions mainly from nonselective backbone cleavages. This behavior was quite different compared to that noted for the prior three proteins (especially for IRMPD), which generally resulted in highly selective fragmentation patterns (i.e., high propensity for backbone cleavages N-terminal to proline and C-terminal to acidic residues). The apparent lower cleavage selectivity for cytochrome C upon IRMPD is attributed to the generation of mostly smaller, singly charged product ions, as discussed previously, in lieu of those arising from proline- and acidic residue-directed cleavages. The smaller ions resulting from secondary dissociation would only show elevated backbone cleavage selectivity if high abundances of proline and acidic residues were located near the N-terminus and C-terminus for a given protein, which for cytochrome C is not the case. We speculate that the enhanced secondary dissociation of cytochrome c is directly related to the presence of the heme group. To support this hypothesis about the special role of IR-absorbing modifications and the impact on secondary dissociation, a single pressure linear quadrupole ion trap (LTQ XL) was employed for the dissociation of the phosphorylated protein α-SI-casein, which has eight, highly IR-absorbing phosphorylation sites. Over 52% of the product ion current

generated upon IRMPD of the 15+ charge state of this protein was attributed to singly and doubly charged product ions (data not shown), a distribution that mirrors the IRMPD results obtained for the heme-containing cytochrome C in which most of the product ions were detected in the lowest charge states. Interestingly, for the case of α-casein, the small, lower charged product ions originated preferentially from selective backbone cleavages N-terminal to proline and C-terminal to acidic residues. The sequence of α-SI-casein protein chain, unlike cytochrome C, has a high frequency of proline and acidic amino acids near the N-terminus and C-terminus.

For all proteins analyzed, IRMPD and CID yielded poor overall sequence coverage - instead fragmentation was directed in both methods to selective backbone cleavages (i.e., N-terminal to proline and C-terminal to acidic residues). Therefore, there were often multiple product ions that resulted from the same backbone cleavage site (i.e., the same products occurring in multiple charge states and/or sequence ions (e.g., a, b, and y) observed for the same backbone cleavage). Regardless, CID in general produced more unique backbone cleavages as compared to IRMPD. Upon CID, ubiquitin (9+), myoglobin (14+), carbonic anhydrase (30+), and cytochrome C (10+) generated product ions originating from 24%, 14%, 12%, and 33% (respectively) of the total possible cleavage sites for each protein. In contrast, IRMPD of ubiquitin (9+, 10+, 12+), myoglobin (16+), carbonic anhydrase (22+), and cytochrome C (16+) produced ions arising from 19%, 9%, 10%, and 29% (respectively) of the total possible cleavage sites for each protein. These particular charge states were chosen for each activation method because they produced the most unique cleavages out of all charge states analyzed for each protein and dissociation method. In general, protein ions in charge states exhibiting the greatest selectivity for cleavages N-terminal to proline and C-terminal to acidic residues ultimately yielded low sequence coverages, as expected.

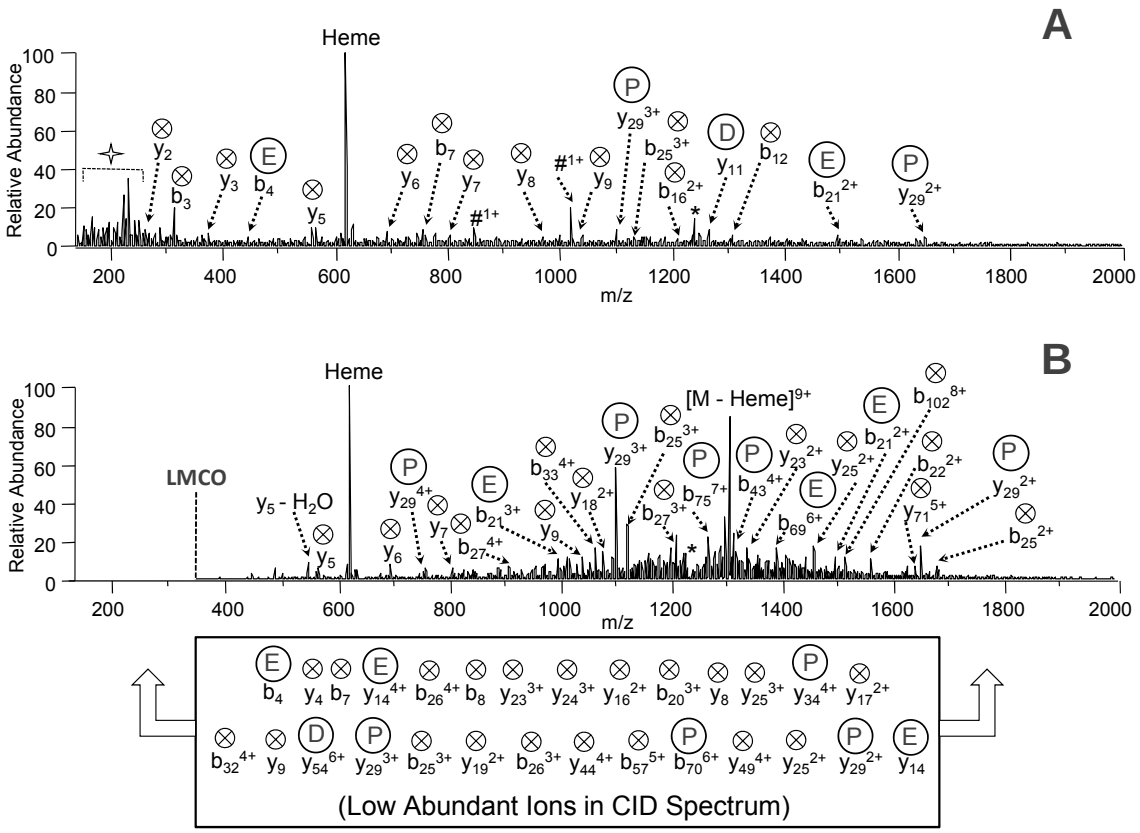

Figure 5.5 Cytochrome C, MW 12,360 Da $[M + 10H]^{10+}$ (*m/z* 1237.0) dissociated by (a) IRMPD in the low pressure cell, t = 11.0 ms and (b) CID in the high pressure cell, 26% NCE. Product ions resulting from selective cleavages N-terminal to proline and C-terminal to acidic residues are denoted by P and D, E (respectively). Internal ions resulting from these selective cleavages are denoted by #, and small internal ions of two amino acids are denoted by ✧. Product ions resulting from unselective cleavages are denoted by ⊗, and precursor ions by *. Due to space constraints in (b), low abundant product ions are listed below the CID spectrum.

5.4.3 IRMPD and CID Product Ion Charge State Comparison

To display the charge state distribution of ions resulting from IRMPD versus those from CID, the peak areas for each observed sequence ion in the various charge states were summed and plotted for all precursor ion charge states of all proteins

precursors (**Figure 5.6**). Percentages are based on the summed peak areas for a particular product ion charge state divided by the total peak areas for all charge states in a given spectrum, and percentage ranges (i.e. 0-20%, 20-40%, etc.) are represented by different colors. To create a more detailed graduated color scale, darker shades of the same color represent higher elevations and thus percentages on the higher end of the range. The distributions shown in **Figure 5.6** illustrate that IRMPD generated product ions in lower charge states relative to CID. In fact, the production of sequence ions in low charges states by IRMPD facilitated their confident assignment, as reflected in **Figure 5.6** by the lack of any product ions in ambiguous charge states in contrast to the companion CID results. The mass spectrometer allows charge states up to 10+ to be isotopically resolved; moreover, most product ions with charge states between 11+ and 16+ could be identified based on their *m/z* values and by limiting the charge states to a few reasonable values based on some resolution of the isotope peaks. In contrast to the IRMPD results, every protein except for the smallest one, ubiquitin, generated significant portions of product ions in high charge states upon CID that could neither be resolved nor identified. This occurrence was especially problematic for the highest precursor charge states. CID also more often generated multiple isobaric product ions that could not be identified due to unresolved, overlapping isotopic patterns of two or more ions, as illustrated in **Figure 5.7** for a product ion (y_{48}^{4+}, *m/z* 1393) of carbonic anhydrase (20+) observed upon IRMPD and CID. Only in the IRMPD mass spectrum could the product ion be isotopically resolved and thus identified as y_{48}^{4+}, whereas the peak in the CID mass spectrum may be the overlap of multiple product ions. Both **Figure 5.6** and **Figure 5.7** illustrate how the production of lower charged sequence ions by IRMPD is particularly beneficial for top-down applications using mass spectrometers which have intrinsically lower resolving powers and mass accuracies, such as quadrupole ion traps. Note that the higher resolution

scan rate (Ultrazoom mode) was not utilized in this study, but could possibly help to resolve the highly charged product ions, or overlapping ions, in some cases.

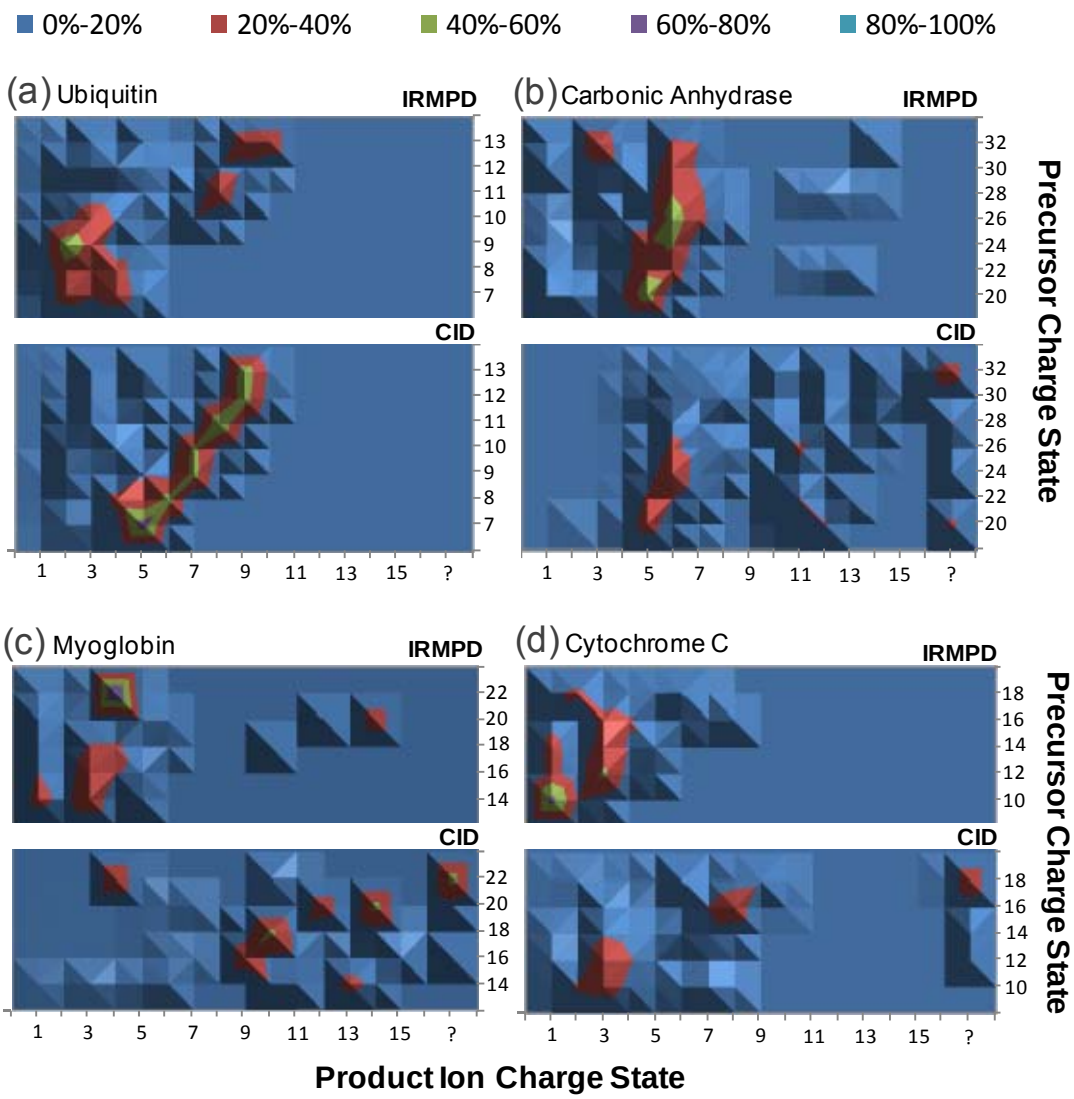

Figure 5.6 Comparison of the charge state distributions of product ions generated by IRMPD versus CID for (a) ubiquitin, (b) carbonic anhydrase, (c) myoglobin, and (d) cytochrome c as a function of precursor charge state. The colored contour plots display the percentages of product ions in specific charge states for each selected precursor charge state. Darker shades of the same color represent higher elevations and correspond to percentages on the higher end of the range. Product ions with ambiguous charge states due to unresolved isotopic profiles are denoted by ?.

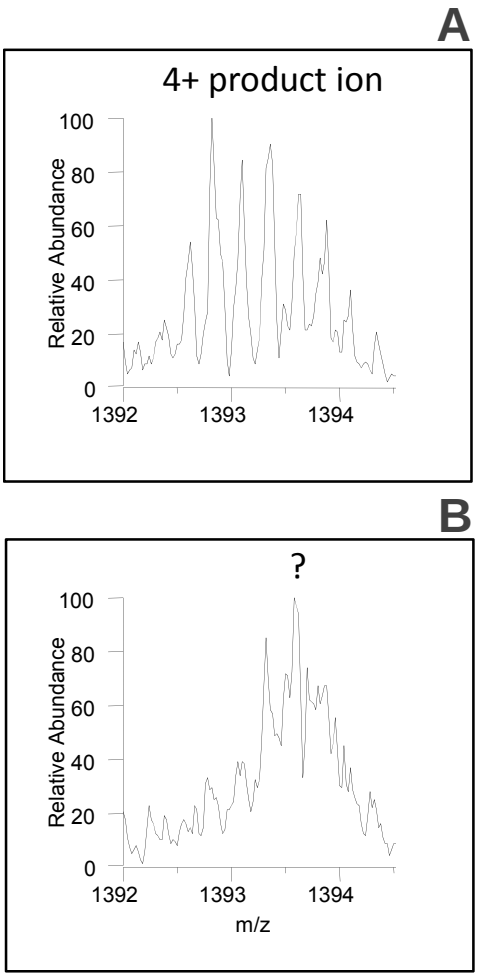

Figure 5.7 (a) Resolved 4+ product ion (*m/z* 1393) after IRMPD of carbonic anhydrase (20+) (b) Unresolved product ion (*m/z* 1393) ion after CID of carbonic anhydrase (20+).

5.4.4 IRMPD and CID Cleavage Selectivity Comparison

The ability to better predict top-down spectra a priori could enhance the design and implementation of algorithms for protein identification. **Figure 5.8** shows a graphical comparison of the distributions of the products created from the most selective backbone cleavages (N-terminal to proline and C-terminal to acidic residues) and those stemming from nonselective backbone cleavages (i.e., all other backbone cleavages that lead to

other a, b and y ions) for all charge states of all proteins by IRMPD and CID. Selective cleavages are represented collectively by different shades of blue, and nonselective backbone cleavages by red. The black boxes that surround individual pie graphs are used to highlight those charge states that yield the greatest portion of selective backbone cleavages (i.e., largest blue segments), both for IRMPD and CID. As seen in **Figure 5.8**, the precursor charge state exerts a significant influence on the backbone cleavage trends for both activation methods, an observation which has been reported previously for CID data.[18, 39] Aside from the results for cytochrome C, the lowest precursor charge states consistently yielded the most selective backbone cleavages for IRMPD, thus indicating that IRMPD of ions in lower charge states offers the best option for producing the most predictable cleavage sites a priori. The decreased cleavage selectivity for the lowest charge state (10+) of cytochrome C upon IRMPD is attributed to the generation of mostly singly charged product ions near the termini of the protein, which for cytochrome C was low in proline and acidic residues, as discussed previously. For CID, there was no predictable pattern with respect to the charge states that showed the greatest backbone cleavage selectivity, and the overall backbone cleavage selectivity of CID never rivaled that obtained by IRMPD (i.e., less substantial blue segments for the CID data in **Figure 5.8**). The enhanced cleavage site selectivity for IRMPD is hypothesized to evolve from the efficient secondary dissociation of large, highly charged product ions initially produced from nonselective backbone cleavage sites (i.e., ones not involving proline or acidic residues). To investigate this phenomenon further, MS^3 experiments using CID were performed in a single pressure linear quadrupole ion trap (LTQ XL) to mimic the type of secondary dissociation that is presumed to occur upon IRMPD in the dual pressure linear ion trap (**Figure 5.9**). Ubiquitin (7+) was dissociated by CID, and the two most abundant product ions resulting from nonselective backbone cleavages (e.g., y_{59}^{5+}

and y_{61}^{5+}) were then individually isolated and subjected to another stage of CID. Interestingly, both of these nonselective products dissociate nearly exclusively into ions resulting from selective backbone cleavages (i.e., ones labeled with P, D, or E in **Figure 5.9**), all of which are similarly observed in the IRMPD spectrum of ubiquitin (7+) obtained in the dual pressure linear ion trap (i.e., compare **Figure 5.2A** with **Figure 5.9**). Therefore, it is evident that secondary dissociation by IRMPD not only reduces the overall charge states of product ions but also ultimately induces greater backbone cleavage selectivity.

Lastly, when using the dual pressure linear ion trap mass spectrometer, the most abundant product ions in the IRMPD spectra obtained for the lowest precursor charge states were attributed invariably to selective backbone cleavages and were also always y ions, which is consistent with previous studies showing y-type ions are more stable than b ions upon IR irradiation.[51] Using this information and the intact molecular weight, it may be possible to more rapidly identify proteins with high confidence by only a few pieces of data – if the protein sequence is available in the database employed.

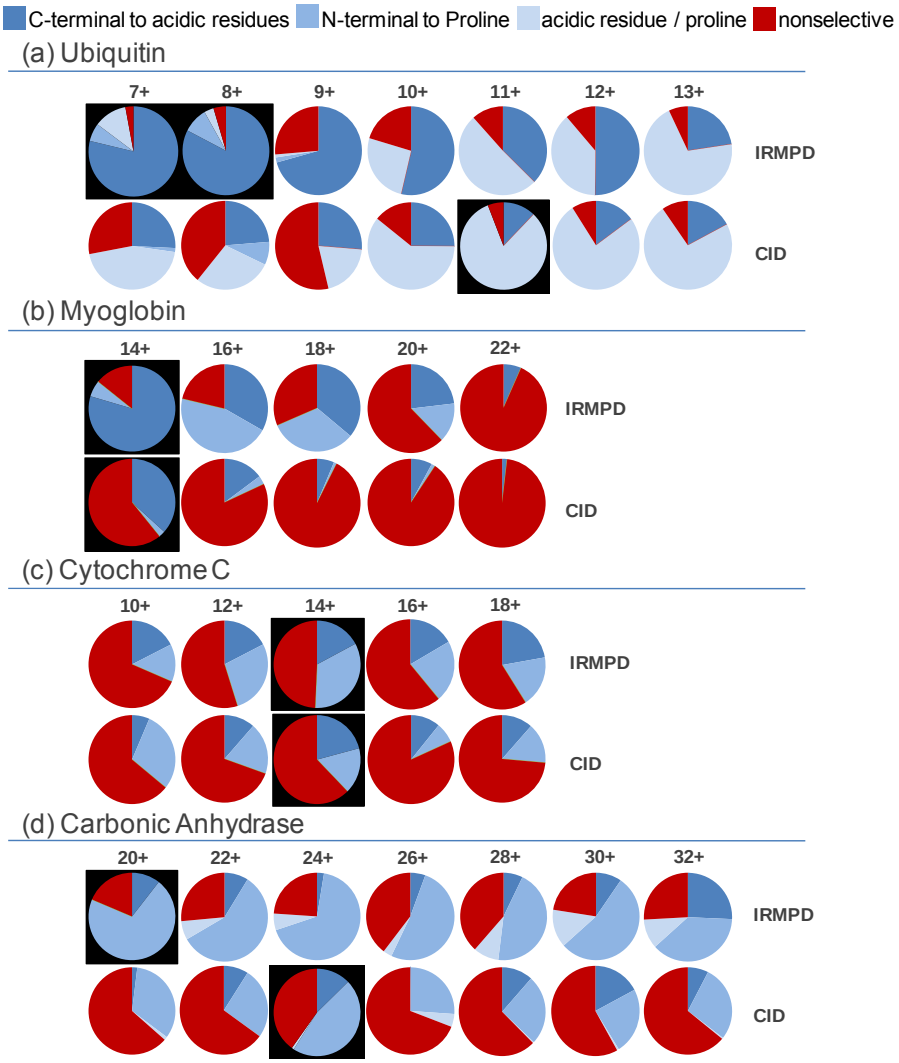

Figure 5.8 Backbone cleavage site comparison of IRMPD versus CID for (a) ubiquitin, (b) myoglobin, (c) cytochrome c, and (d) carbonic anhydrase. The blue-shaded areas represent the portion of product ions that stem from backbone cleavages that occur C-terminal to acidic residues and/or N-terminal to proline residues. Those product ions that arise from cleavages between DP or EP residues may be classified as both N-terminal to proline and C-terminal to acidic residues and thus are collectively categorized as "acidic residue/proline" ions (lightest blue color). The red areas represent the portion of products arising from other nonselective backbone cleavages. The black boxes highlight those charge states that dissociated (by IRMPD or CID with the highest overall selectivity towards backbone cleavages that occur C-terminal to acidic residues and/or N-terminal to proline.

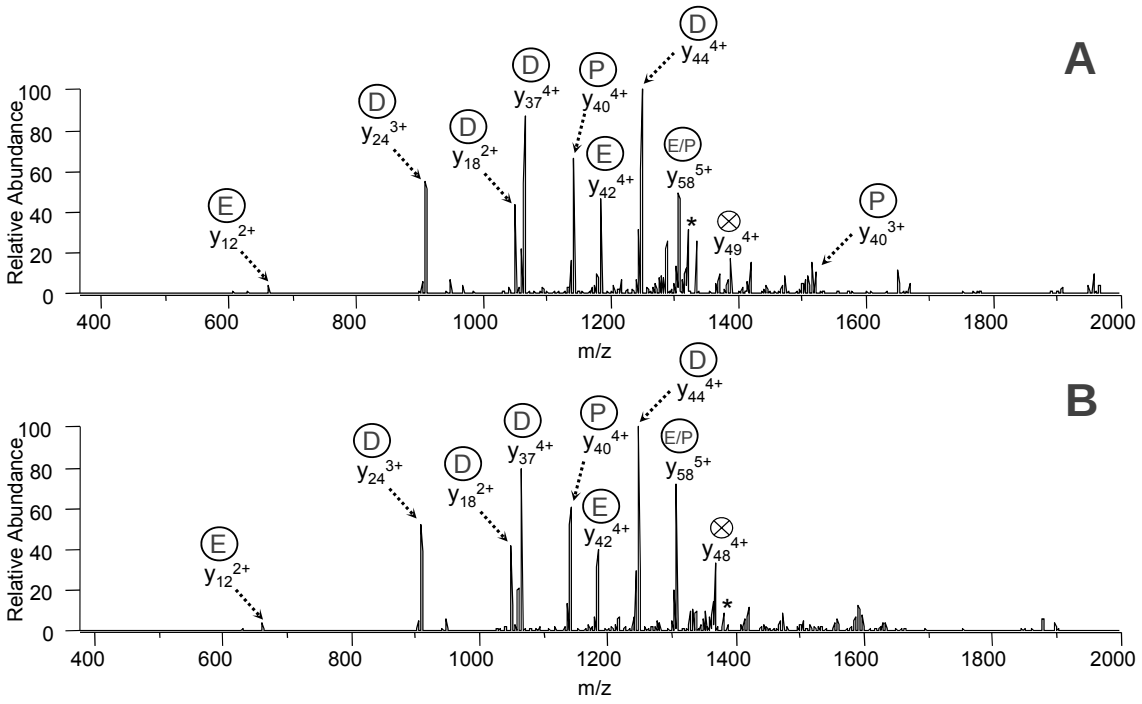

Figure 5.9 MS^3 of ions produced from unselective backbone cleavages of ubiquitin. (a) CID of $[M + 7H]^{7+}$ of *m/z* 1224.5(15% NCE) followed by CID of the y_{59}^{5+} product ion of *m/z* 1332.5 (19% NCE). (b) CID of $[M + 7H]^{7+}$ of *m/z* 1224.5 (15% NCE) followed by CID of the y_{61}^{5+} product ion of *m/z* 1378.5 (19% NCE). Product ions resulting from selective cleavages N-terminal to proline and C-terminal to acidic residues are denoted by P and D, E (respectively). Product ions resulting from unselective cleavages are denoted by ⊗, and precursor ions by *.

5.4.5 Current Top-down Algorithm Searching

Prosight PTM 2.0 was employed to investigate the use of one of the current top-down searching algorithms for protein identification based on the raw data from the dual pressure linear ion trap IRMPD and CID methods. For each *Bos taurus* protein (e.g., ubiquitin and carbonic anhydrase), the charge states that yielded the most selective backbone cleavages and the charge states with the greatest sequence coverage (i.e., the greatest number of unique backbone cleavages) for both IRMPD and CID were selected

for ProsightPTM searching against the *Bos taurus* protein database **(Table 5.1**). For ubiquitin, the top ranked protein identified from the database search had the correct sequence for both dissociation methods and each charge state (i.e., most selective and greatest sequence coverage) when considering only the probability and expectation scores. IRMPD yielded similar and satisfactory scores for each charge state (i.e. lower p-scores and lower expectation values signify more confident matches). However, the CID scores were quite varied; the charge state (11+) that exhibited the most selective cleavage behavior yielded the worst score of any ubiquitin charge state (e.g., p-score = 2.25×10^{-6}), and the charge state giving the greatest sequence coverage (9+) had the best score of all charge states (e.g., p-score = 2.16×10^{-18}) among both the CID and IRMPD data. Conversely, for the much larger and N-terminally modified (acetylation) carbonic anhydrase protein, IRMPD outperformed CID most likely due to the more accurate mass assignments of the inherently lower charged product ions. Again, the protein was successfully identified via the IRMPD data, and both charge states (20+, 22+) yielded similar and satisfactory scores. The CID data on the other hand was less successful for identification of the protein (i.e., the correct sequence was ranked third) and yielded poor p-scores for both charge states (p-score of 2.60×10^{-5} for the 24+ charge state and p-score 1.42×10^{-7} for the 30+ charge state). Since smaller proteins such as ubiquitin have lower precursor charge states compared to larger proteins, product ions are generated in relatively low charge states which allow more accurate mass assignments of the fragment ions produced by either method (CID or IRMPD) in the dual pressure quadrupole linear ion trap. As a result, CID may be the best option for small protein identification using current top-down searching algorithms (based on probability and expectation scoring) since higher sequence coverage is generally achieved. IRMPD, however, offers clear advantages for larger proteins as earlier discussed. The charge state also played an

important role in the success of CID; however, a large difference was not found in the scores generated from IRMPD mass spectra of proteins in the charge states that yielded the most selective cleavages versus those that yielded the greatest number of sequence ions. Interestingly, even though IRMPD consistently produced fewer sequence ions and resulted in lower sequence coverage as compared to CID, protein identification was more successful by IRMPD because product ions were assigned with greater confidence due to their lower charge states.

Since IRMPD yields abundant product ions arising from cleavage sites adjacent to N-terminal prolines and C-terminal acidic residues (glutamic acid and aspartic acid), PDE scores, which take into account these selective cleavages, were also used to gauge protein identification confidence (**Table 5.1**). PDE scoring utilizes weighting factors that are multiplied by the abundances of product ions that result from selective backbone cleavages as described in the Experimental Section of this study. The PDE weighting factors include 7x for ions arising from cleavages N-terminal to proline and C-terminal to glutamic acid (EP), 5x for ions related to cleavages N-terminal to proline (P), 5x for ions stemming from cleavages C-terminal to aspartic acid (D), and 2x for ions arising from C-terminal to glutamic acid (E) cleavages. As expected, the charge states that promoted the most selective fragmentation (for both activation methods) resulted in the highest and best scores even with inherently lower total numbers of product ions and lower sequence coverage. For ubiquitin, the highest PDE score (560) was obtained from the CID mass spectrum of the 11+ charge state due to a very high percentage of product ion signal attributed to cleavages between glutamic acid and proline (EP) residues. However, the charge state (7+) that exhibited the highest backbone cleavage selectivity upon IRMPD also had a high PDE score (384), but unlike CID, was in a predictable low charge state. For the larger carbonic anhydrase protein, the charge state that yielded the greatest

cleavage selectivity upon IRMPD (20+) gave the highest PDE score as compared to the charge state that yielded the greatest sequence coverage by IRMPD (22+) or each charge state by CID (i.e., the ones which produced the most selective and the greatest sequence coverage).

In summary, IRMPD in the dual pressure linear ion trap allowed confident protein identification using various scoring options in top-down searching algorithms. Due to the distinctive and elevated backbone cleavage selectivity observed upon IRMPD over CID, protein identification confidence potentially could be increased further by placing an even higher emphasis on backbone cleavages N-terminal to proline and C-terminal to acidic residues, especially when undertaking IRMPD on the lowest precursor charge states.

Protein (charge state)	dissociation method*	dissociation behavior	ranking	p-score	expectation	PDE
Ubiquitin (7+)	IRMPD	most selective	1	6.08×10^{-10}	1.74×10^{-3}	384
Ubiquitin (9+)	IRMPD	most coverage	1	3.29×10^{-9}	9.43×10^{-3}	189
Ubiquitin (11+)	CID	most selective	1	2.25×10^{-6}	6.45	560
Ubiquitin (9+)	CID	most coverage	1	2.16×10^{-18}	6.19×10^{-12}	105
Carbonic Anhydrase (20+)	IRMPD	most selective	1	3.65×10^{-11}	1.05×10^{-4}	270
Carbonic Anhydrase (22+)	IRMPD	most coverage	1	2.72×10^{-10}	7.81×10^{-4}	228
Carbonic Anhydrase (24+)	CID	most selective	3	2.60×10^{-5}	74.7	157
Carbonic Anhydrase (30+)	CID	most coverage	3	1.42×10^{-7}	4.07×10^{-1}	143

*IRMPD was performed in the low pressure cell; CID was performed in the high pressure cell.

Table 5.1 ProsightPTM Database search of *Bos taurus* proteins

5.5 Conclusions

IRMPD afforded efficient dissociation of intact protein cations in a dual pressure linear ion trap. Product ions from IRMPD were formed in significantly lower charge states compared to the ones generated by CID. Furthermore, IRMPD of the lowest charge states of the protein precursor ions produced the highest degree of selective backbone cleavages (i.e., N-terminal to proline and C-terminal to acidic residues) as compared to higher charge states of the same method or any charge state by CID. These attributes of IRMPD are particularly beneficial for a priori spectral predictions and improved algorithm searching for protein identification by top-down methods using trapping instruments. Furthermore, with the use of ion parking via ion-ion reactions prior to ion activation,[50, 52] the ion signal could be concentrated in a low precursor charge state followed by IRMPD to yield the highest backbone cleavage selectivity with maximum sensitivity.

5.6 References

(1) Chait, B. T. *Science* **2006**, *314*, 65-66.
(2) Collier, T. S.; Hawkridge, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C. *Analytical Chemistry* **2008**, *80*, 4994-5001.
(3) Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.
(4) Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne, M. T., II; Burke, P. V.; Kwast, K. E.; Kelleher, N. L. *Analytical Chemistry* **2007**, *79*, 7984-7991.
(5) Liu, J.; Huang, T.-Y.; McLuckey, S. A. *Analytical Chemistry* **2009**, *81*, 1433-1441.
(6) Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M. W.; McLafferty, F. W. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 286-289.
(7) Senko, M. W.; Speir, J. P.; McLafferty, F. W. *Analytical Chemistry* **1994**, *66*, 2801-2808.
(8) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *Journal of the American Chemical Society* **1999**, *121*, 806-812.
(9) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.
(10) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Analytical Chemistry* **2000**, *72*, 563-573.
(11) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 10313-10317.
(12) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Analytical Chemistry* **1994**, *66*, 2809-2815.
(13) Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little Daniel, P.; McLafferty, F. W. *International Journal of Mass Spectrometry and Ion Processes* **1996**, *157/158*, 357-364.
(14) Liu, J.; Huang, T.-Y.; McLuckey, S. A. *Analytical Chemistry* **2009**, *81*, 2159-2167.
(15) Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. *Molecular and Cellular Proteomics* **2006**, *5*, 949-958.
(16) Amunugama, R.; Hogan, J. M.; Newton, K. A.; McLuckey, S. A. *Analytical Chemistry* **2004**, *76*, 720-727.
(17) Hogan, J. M.; Pitteri, S. J.; McLuckey, S. A. *Analytical Chemistry* **2003**, *75*, 6509-6516.
(18) Reid, G. E.; McLuckey, S. A. *Journal of Mass Spectrometry* **2002**, *37*, 663-675.
(19) Bowers, J. J.; Liu, J.; Gunawardena, H. P.; McLuckey, S. A. *Journal of Mass Spectrometry* **2008**, *43*, 23-34.

(20) Coon, J. J.; Ueberheide, B.; Syka, J. E. P.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 9463-9468.
(21) Bunger, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L., Jr. *Analytical Chemistry* **2008**, *80*, 1459-1467.
(22) Colorado, A.; Shen, J. X.; Vartanian, V. H.; Brodbelt, J. *Analytical Chemistry* **1996**, *68*, 4033-4043.
(23) Crowe, M. C.; Brodbelt, J. S. *Analytical Chemistry* **2005**, *77*, 5726-5734.
(24) Madsen, J. A.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 349-358.
(25) Crowe, M. C.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 1581-1592.
(26) Flora, J. W.; Muddiman, D. C. *Analytical Chemistry* **2001**, *73*, 3305-3311.
(27) Flora, J. W.; Muddiman, D. C. *Journal of the American Chemical Society* **2002**, *124*, 6546-6547.
(28) Flora, J. W.; Muddiman, D. C. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 121-127.
(29) Pikulski, M.; Wilson, J. J.; Aguilar, A.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 8512-8517.
(30) Pikulski, M.; Hargrove, A.; Shabbir, S. H.; Anslyn, E. V.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 2094-2106.
(31) Wilson, J. J.; Brodbelt, J. S. *Analytical Chemistry* **2006**, *78*, 6855-6862.
(32) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80*, 4807-4819.
(33) Newsome, G. A.; Glish, G. L. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1127-1131.
(34) Payne, A. H.; Glish, G. L. *Analytical Chemistry* **2001**, *73*, 3542-3548.
(35) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C., Jr.; Brodbelt, J. S. *Analytical Chemistry* **2009**, submitted.
(36) Raspopov, S. A.; El-Faramawy, A.; Thomson, B. A.; Siu, K. W. M. *Analytical Chemistry* **2006**, *78*, 4572-4577.
(37) Scherperel, G.; Reid, G. E. *Analyst* **2007**, *132*, 500-506.
(38) Ly, T.; Julian, R. R. *Journal of the American Chemical Society* **2008**, *130*, 351-358.
(39) Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. *Analytical Chemistry* **2001**, *73*, 3274-3281.
(40) Watson, D. J.; McLuckey, S. A. *International Journal of Mass Spectrometry* **2006**, *255-256*, 53-64.
(41) Iavarone, A. T.; Williams, E. R. *Analytical Chemistry* **2003**, *75*, 4525-4533.
(42) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Analytical Chemistry* **1993**, *65*, 425-438.

(43) Bakhtiar, R.; Wu, Q.; Hofstadler, S. A.; Smith, R. D. *Biological Mass Spectrometry* **1994**, *23*, 707-710.
(44) Schwartz, J. C.; Syka, J. E. P.; Remes, P. M. *Proceedings of the 56th ASMS Conference on Mass Spectrometry and Allied Topics* **2008**, Denver, CO, June 2008.
(45) Taylor, G. K.; Kim, Y.-B.; Forbes, A. J.; Meng, F.; McCarthy, R.; Kelleher, N. L. *Analytical Chemistry* **2003**, *75*, 4081-4086.
(46) LeDuc, R. D.; Taylor, G. K.; Kim, Y.-B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. *Nucleic Acids Research* **2004**, *32*, W340-W345.
(47) Zamdborg, L.; LeDuc Richard, D.; Glowacz Kevin, J.; Kim, Y.-B.; Viswanathan, V.; Spaulding Ian, T.; Early Bryan, P.; Bluhm Eric, J.; Babai, S.; Kelleher Neil, L. *Nucleic Acids Research* **2007**, *35*, W701-706.
(48) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nature Biotechnology* **2001**, *19*, 952-957.
(49) Sadygov, R. G.; Yates, J. R. *Analytical Chemistry* **2003**, *75*, 3792-3798.
(50) Reid, G. E.; Shang, H.; Hogan, J. M.; Lee, G. U.; McLuckey, S. A. *Journal of the American Chemical Society* **2002**, *124*, 7353-7362.
(51) Aaserud, D. J.; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. *Rapid Communications in Mass Spectrometry* **1995**, *9*, 871-876.
(52) McLuckey, S. A.; Reid, G. E.; Wells, J. M. *Analytical Chemistry* **2002**, *74*, 336-346.

Chapter 6

Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflow

6.1 Overview

Ultraviolet photodissociation (UVPD) at 193 nm was implemented on a linear ion trap mass spectrometer for high-throughput proteomic workflows. Upon irradiation by a single 5 ns laser pulse, efficient photodissociation of tryptic peptides was achieved with production of *a*, *b*, *c*, *x*, *y*, and *z* sequence ions, in addition to immonium ions and *v* and *w* side-chain loss ions. The factors that influence the UVPD mass spectra and subsequent *in silico* database searching via SEQUEST were evaluated. Peptide sequence aromaticity and the precursor charge state were found to influence photodissociation efficiency more so than the number of amide chromophores, and the ion trap *q*-value and number of laser pulses significantly affected the number and abundances of diagnostic product ions (e.g., sequence and immonium ions). Also, photoionization background subtraction was shown to dramatically improve SEQUEST results, especially when peptide signals were low. A liquid chromatography – mass spectrometry (LC-MS) – UVPD strategy was implemented and yielded comparable or better results relative to LC-MS – collision induced dissociation (CID) for analysis of proteolyzed bovine serum albumin and lysed human HT-1080 cytosolic fibrosarcoma cells.

6.2 Introduction

Large scale, "bottom-up" characterization of cellular proteomes has been enormously successful[1] and continues to improve due to ongoing innovation in the areas of mass spectrometric instrumentation, tandem mass spectrometry (MS^n) techniques, sampling, separations, and bioinformatics. Typical high-throughput, bottom-up workflows consist of the separation of complex mixtures of digested proteins followed by online mass spectrometry (MS) and MS^n analysis. Proteins are then qualitatively and/or quantitatively identified by *in silico* database searching algorithms that interpret the MS data. Four of the most popular algorithms, SEQUEST,[2] MASCOT,[3] OMSSA,[4] and X! Tandem,[5] were recently compared,[6] and each algorithm yielded acceptable and similar results for a complex human protein sample. In terms of MS instrumentation, the introduction of the Orbitrap[7, 8] and hybrid linear ion trap (LIT)-Orbitraps[9, 10] have afforded significantly better mass accuracy and resolution (i.e., a resolving power of ~80,000 – 100,000 at *m/z* of around 1000)[7], which has greatly increased selectivity in database searching for bottom-up experiments. Also, the analysis times of mass spectrometers continue to decrease. The newly released dual-pressure linear ion trap (Velos),[11] for example, has decreased cycle times two-fold by technological advances that eliminate prescans and allow faster scan rates in a low pressure trap. This improvement significantly increased the experimental duty cycle over more traditional trapping instruments, resulting in more protein identifications and a deeper depth of analysis into the proteome.[11] Moreover, the benefits and growing popularity of ultrafast separation techniques (e.g., ultra high performance liquid chromatography (UHPLC), microfluidics, capillary electrophoresis (CE), etc.), will certainly spur the development of even faster MS instrumentation as well as strategies that can efficiently sample extremely narrow peaks.

For successful protein identification, bottom-up approaches rely on the collection of informative tandem mass spectra that critically depend on the activation technique used. In recent years there have been many advances in activation technology, but there still remains no universal method that can successfully and efficiently analyze all proteins and their peptide constituents. The two most popular tandem mass spectrometric techniques used in large scale analysis of cellular proteins are the traditional collision induced dissociation (CID) and the more recently introduced electron transfer dissociation (ETD).[12] The latter has shown the most success for identifying and sequencing phosphorylated peptides and was recently shown to significantly outperform CID in characterizing the phosphoproteome of various human cells.[13, 14] CID on the other hand, is an established, sensitive, and robust method that has successfully identified the most proteins based on analysis of complex tryptic peptide mixtures in shotgun bottom-up experiments.[15, 16] Higher-energy C-trap dissociation (HCD), a form of CID that allows smaller *m/z* product ions to be detected, has recently been implemented on the newest hybrid Orbitraps and has yielded comparable results to that of CID for complex biological samples.[10, 17] All of the techniques just described (CID, ETD, and HCD) require between 10 – 100 ms of activation time (depending on the specific instrument and analytes) for efficient precursor dissociation, posing a significant drawback with respect to implementation of high throughput strategies.

The use of single or multiple photons for activation and dissociation of peptides has also shown promise. [18-21] Ultraviolet photodissociation at 193 nm of peptides was first performed in the 1980's,[22-25] but experiments were limited to a few selected peptides on Fourier transform ion cyclotron resonance (FTICR) mass spectrometers. A recent renaissance in the use of short wavelength (157 and 193 nm)[26-31] ultraviolet photodissociation (UVPD) coupled mainly to time-of-flight (TOF) instruments has

occurred due to the ultrafast (nanosecond) activation timescales and the rich tandem mass spectrometric information obtained (i.e., more types of backbone ions are observed compared to traditional activation methods). There have been noteworthy inroads in the implementation of MALDI-TOF-MS with UVPD for high-throughput proteomic workflows,[32] although inefficient ion activation has hindered the analysis of the lowest abundance proteins/peptides,[27] coupled with the typical need for ~2000 spectral averages (~40 s) for each spectrum.[26, 32] Another recently introduced and promising photodissociation method, femtosecond laser induced ionization/dissociation (fs-LID), produced rich fragmentation patterns for peptides (i.e., *a*, *b*, *c*, *x*, *y*, and *z* ions were all observed) using extremely fast femtosecond laser pulses, yet required rather long total activation periods (up to 200 ms).[33]

If a single short UV pulse (5 ns) was employed for photodissociation for high-throughput liquid chromatography – tandem mass spectrometry (LC-MS/MS) instead of using conventional CID (30 msec activation time), an additional 1800 peaks could conceivably be analyzed per hour of analysis using traditional quadrupole ion traps.[20] This increase in duty cycle would be even more dramatic when using newer generation linear ion traps,[11] or when directly compared to strategies that use the slower activation methods such as ETD (typically 100 msec activation time). However, if the spectral information produced by UVPD pales in comparison to that afforded by gold standard methods such as CID, then the benefits of ultrafast photoactivation will be negated. Therefore, a thorough comparison of UVPD and CID as applied to LC-MS/MS analysis and *in silico* data interpretation is warranted.

In the present study, UVPD at 193 nm is implemented on a linear ion trap mass spectrometer for high-throughput proteomic workflows. Efficient photodissociation of tryptic peptides is achieved using a single 5 ns laser pulse with minimal need for spectral

averaging. The factors that affect the optimization of universal photodissociation parameters for LC-MS/UVPD experiments and successful *in silico* database searching via SEQUEST are investigated. This ultrafast photodissociation method yields comparable or better results compared to CID for complex samples of proteolyzed bovine serum albumin (BSA) (1 μg) and lysed human HT-1080 cytosolic fibrosarcoma cells (10 μg lysate) and represents a high-throughput LCMS/MS strategy based on photodissociation and database searching for the analysis of biologically relevant samples.

6.3 Experimental

6.3.1 Materials

All acids and buffer components, BSA, proteomics grade trypsin, and all model peptides were purchased from Sigma-Aldrich (St. Louis, MO), except the peptides KRPPGFSPFR, YRPPGFSPFR, DRVYIHPFHLVIH, DAEFRHDSGYQVHHQK, HCKFWW, GNHWAVGHLM, and ASHLGLAR, which were purchased from BACHEM (King of Prussia, PA). KAKAA and RAAAA were synthesized by Bio Basic (Ontario, Canada). Solvents for direct infusion and HPLC were obtained from Fisher Scientific (Fairlawn, NJ).

6.3.2 Sample Preparation

BSA was reduced by adding a 10-fold molar excess of dithiothreitol (DTT) (100 μL of 1.0 mM DTT) to 10 nmoles of BSA (10 μL of 1.0 mM protein), which was buffered to a pH of ~8.4 with ~4 μmoles ammonium bicarbonate. This solution was then incubated for one hour at 40 ˚C after which alkylation was performed by adding 4.0 μL of 1.0 M iodoacetamide (buffered to a pH of ~8.0 with ~20 μmoles ammonium

bicarbonate). This solution was incubated at ambient temperature for 45 minutes in the dark followed by subsequent quenching by the addition of excess DTT. For enzymatic digestion, 5.0 μL of 1.0 mg/mL trypsin in 1.0 mM HCl was added to the reduced/alkylated solution followed by incubation at 37˚C for 16 hours.

HT1080 cells were suspended in low-salt buffer (10 mM Tris-HCl, 10 mM KCl, 1.5 mM $MgCl_2$, pH 8.0) to swell, and were then lysed by dounce homogenization. The whole cell lysate was centrifuged at 1000 × g to clarify the soluble lysate and to remove the insoluble pellet. Soluble lysate (2 mg/mL protein) was denatured by addition of 2,2,2-trifluoroethanol (TFE) to 50% (vol/vol) and reduced with 15 mM DTT at 55˚C for 45 min. Following reduction, the proteins were alkylated with 55 mM iodoacetamide at room temperature for 30 min. The sample was diluted in digestion buffer (50 mM Tris-HCl, 2 mM $CaCl_2$, pH 8.0) to a final TFE concentration of 5% (vol/vol). Trypsin was added to a 1:50 (enzyme:protein) concentration (w/w), and the sample was incubated at 37˚C for 5 hrs. The digestion was quenched with 1% formic acid and sample volume was reduced to 20 μl by centrifugation under reduced pressure using a SpeedVac. Digested peptides were bound and washed on Thermo Fisher Scientific HyperSep C-18 SpinTips (San Jose, CA), resuspended in peptide buffer (95% H_2O/5% acetonitrile/0.1% formic acid) and filtered through Millipore Microcon 10 kDa centrifugal filters (Billerica, MA), with the digested peptides collected as flow-through.

6.3.3 Mass Spectrometry, Ultraviolet Photodissociation, and Liquid Chromatography

Analysis by mass spectrometry was performed on a Thermo Fisher Scientific LTQ XL linear ion trap (LIT) mass spectrometer (San Jose, CA) outfitted with a Coherent ExciStar XS excimer laser (Santa Clara, CA) operated at 193 nm. The laser

setup was similar to that previously described,[34] except a CaF_2 lens was used to transmit 193 nm photons into the LIT. For UVPD experiments, the laser was pulsed once per scan with an energy of 8 mJ/pulse and a pulse duration of 5 ns. Pulse variable experiments used a laser pulse frequency of 500 Hz. For CID, a 30 ms activation time and a *q*-value of 0.25 were used. For direct infusion, peptides were diluted to 10 μM in 50/50 MeOH/H_2O and 1% acetic acid prior to ESI-MS/MS analysis. The automated gain control (AGC) was set to 3×10^4 for MS and 1×10^4 for MS^n scans. An ESI voltage of 4 kV and a heated capillary temperature of 180°C were used for all mass spectrometry experiments.

Liquid chromatography was performed using a Dionex UltiMate 3000 system (Sunnyvale, CA) using a capillary flow splitter. An Agilent ZORBAX 300SB-C_{18} column (Santa Clara, CA) (150 × 0.3 mm, 5 μm particle size) was used for all separations. The column temperature was kept constant at 40°C. Eluent A consisted of 0.1% formic acid in water and eluent B 0.1% formic acid in acetonitrile. A linear gradient from 5% eluent B to 40% eluent B over 65 min at 5 μL/min was used. Injections of approximately 1 μg (20 picomoles) were used for the digested BSA sample and 10 μg for the HT-1080 lysate sample. Data-dependent LC-MS/MS was performed in two different ways. For LC-MS/UVPD runs (i.e., only UVPD was used for activation), the first event was the full mass scan (*m/z* range of 400 – 2000) followed by ten consecutive UVPD events on the ten most abundant ions from the full mass scan with one UV pulse applied per MS/MS scan. A *q*-value of 0.1 was used for each UVPD event. For LC-MS/UVPD/CID runs (i.e., comparison between UVPD and CID), the first event was the full mass scan (*m/z* range of 400 – 2000) followed by ten alternating UVPD/CID events on the five most abundant peaks for a total of five UVPD and five CID events per cycle. A *q*-value of 0.25, an activation time of 30 ms, and a normalized collision energy (NCE) of 35% were used for each CID scan event; NCE = (voltage × 30) / (0.000062 × *m/z* × 0.006325) in which

0.000062 is the TickAmpSlope and 0.006325 is the TickAmpInt. A *q*-value of 0.1 and a default activation period of 30 μs were used for each UVPD event. Although the actual irradiation time was only 5 ns (the duration of a single laser pulse) for UVPD, the commercial LTQ software limited the activation period to a minimum of 30 μs. The maximum injection time for all mass scan and MS/MS events was set to 100 ms. The dynamic exclusion duration was set to 50 s, the exclusion list size allowed for 500 specified m/z values, and a single repeat count for LC-MS/UVPD and two repeat counts for LC-MS/UVPD/CID experiments. Each centroid mass spectrum and tandem mass spectrum was the average of three microscans. UVPD precursor dissociation efficiencies as well as pulse-variable and *q*-variable plots were calculated from ion peak areas using Origin 7.0. Photodissociation efficiency (expressed as a percentage) is defined as follows: 100 – [(surviving precursor abundance/initial isolated precursor abundance) x 100)], in which the precursor abundances are measured as ion peak areas.

6.3.4 Background Subtraction and Data Analysis

Prior to SEQUEST interpretation of the MS/MS data, UVPD mass spectra were subjected to a background subtraction procedure to reduce/eliminate photoionization products as described in detail in the results and discussion section. Background spectra were collected when no sample was being infused and thus no ions were detected in the trap, and at isolation *m/z* values of 250, 500, 750, or 1000 (isolation width of 4 *m/z*), using a protocol discussed in more detail later. Each centroid background spectra was the culmination of 20 averages of three microscans. The LC-MS/UVPD RAW files were subtracted from the background RAW files using Thermo Fisher Xcalibur version 2.0.7 software.

SEQUEST was used for *in silico* MS/MS interpretation through the Thermo Fisher Scientific Proteome Discoverer 1.0 software package. For SEQUEST, a signal:noise ratio of 3, a precursor mass tolerance of 1.5 Da, and a fragment mass tolerance of 0.8 Da were used. Product ion series for UVPD included *a*, *b*, *c*, *x*, *y*, and *z* ions, and for CID included *b* and *y* ions except where noted. UVPD and CID spectra were manually separated from LC-MS-UVPD/CID files for direct comparison. Non-redundant, bovine (25,243 sequences) and human (33,819 sequences) protein databases from the NCBI were used for searching of BSA and HT-1080 samples, respectively. Oxidation of methionines and carbamidomethyl modification of cysteines were set as dynamic and static (respectively) side chain modifications. Peptide hits were filtered against decoy databases at a 1% false discovery rate (FDR) (based on Xcorr vs. charge), and were manually verified by the presence of matching immonium ions and predicted sequence ions (from selective cleavages). Peptides with less than six amino acids were also filtered out.

6.4 Results and Discussion

Activation of multiply-charged peptide cations upon absorption of 193 nm photons produces a mixture of *a*, *b*, *c*, *x*, *y*, and *z* ions and a few *w* and *v* side-chain loss ions. Foremost the *b*/*y* series followed by the *a*/*x* series are generally observed as the most abundant products, often yielding almost full sequence coverage from each of these four ion types. These spectra will be discussed in more detail later in the context of the study. As previously observed for photodissociation of peptides at 157 nm,[30] singly-charged precursors result in more abundant products arising from side-chain losses as compared to those from multiply-charged precursors, which predominate in LC-MS/MS analysis.

Therefore, the generation of an extensive array of abundant *a*, *b*, *c*, *x*, *y*, and *z* sequence ions in addition to immonium ions indicates that the UVPD spectra at 193 nm are well suited to *in silico* database searching, as shown in this study.

6.4.1 Impact of Aromatic Side-chains and Charge State on UVPD

Having a set of robust and universal activation parameters that will yield both high dissociation efficiencies and informative product ions is important for analyzing a variety of peptides of different sizes, amino acid sequences, and charge states in a single LC-MS/MS run. In the present study, the UV photodissociation efficiencies were evaluated for a series of peptides of varying lengths, charge states, and sequences, including ones containing residues with aromatic side-chains. The results are summarized in bar graph form in **Figure 6.1A**, in which the abundance of the surviving precursor is monitored relative to the abundance of the original precursor ion, thus allowing estimation of photodissociation efficiencies using a single (5 ns, 8 mJ) pulse. **Figure 6.1A** shows the photodissociation efficiencies relative to the number of aromatic side-chains, ranging from zero to three. The photodissociation efficiency increases with the number of amino acids containing aromatic side-chains (tryptophan, tyrosine, phenylalanine), with efficiencies ranging from 50% for ASHLGLAR (2+) to 98% for HCKFWW (2+). This trend was also observed for peptides with nearly identical sequences (e.g., KRPPGFSPFR, RPPGFSPFR, and YRPPGFSPFR, all 2+), but differing by one aromatic group. KRPPGFSPFR and RPPGFSPFR each contained two aromatic groups and yielded similar photodissociation efficiencies of 80 $\pm$ 2% and 78 $\pm$ 3%, respectively. YRPPGFSPFR, on the other hand, contains an additional tyrosine and resulted in a dissociation efficiency of 92 $\pm$ 2%.

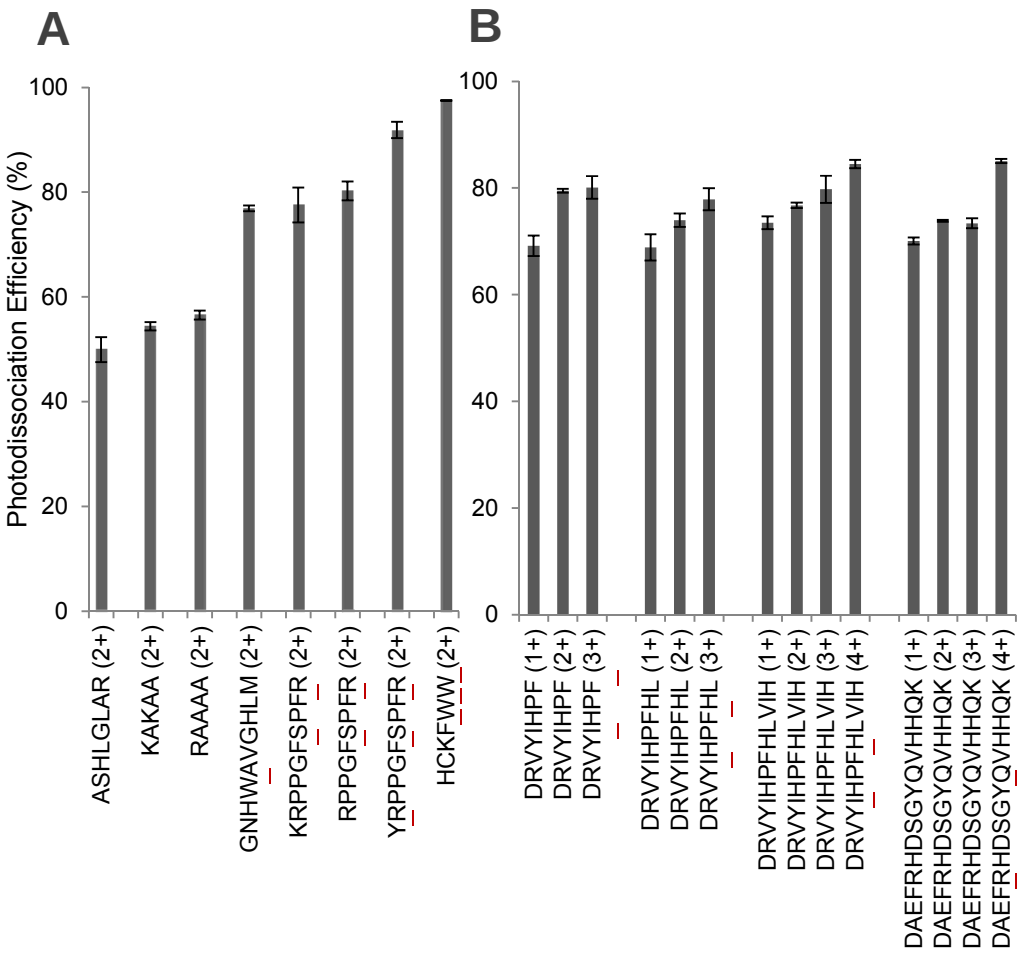

Figure 6.1 (a) Photodissociation efficiencies of peptides (2+ charge state) with varying numbers of aromatic side-chain residues (phenylalanine, tyrosine, and tryptophan); the number of aromatic side-chains increases from left to right. (b) Photodissociation efficiencies of peptides with constant aromatic side-chain residues (one phenylalanine and one tyrosine) with varying peptide length and charge state. Each peptide was dissociated with one 5 ns, 8mJ pulse and a q value of 0.1. Photodissociation efficiency is defined as follows: 100 – [(surviving precursor abundance/isolated precursor abundance) × 100). Aromatic residues are underlined.

The total abundances of the immonium product ions (based on peak areas) were plotted as a function of the number of laser pulses for the doubly-charged YRPPGFSPFR (**Figure 6.2**). The abundances of the arginine (R), serine (S), and proline (P) immonium

ions increase on going from one to two UV pulses and plateau after two laser pulses. However, for the phenylalanine (F) and tyrosine (Y) immonium ions, the abundances of the Y immonium ions steadily decrease after the first pulse and likewise the abundances of the F ions diminish after the third pulse (see inset), a result consistent with their enhanced secondary dissociation due to higher photoabsorptivity. Similarly, the decrease in the abundance of the tyrosine immonium ions substantially exceeded the decrease in abundance of the phenylalanine immonium ions, suggesting a significant difference in photoabsorptivities or stabilities of these two species. The general consensus for vacuum UVPD (157 and 193 nm) is that photodissociation efficiency is generally mediated by the photoabsorptivities of the backbone amide chromophores.[19, 24, 31] While this is true for peptides with no F, Y, and W residues, the presence of amino acids bearing these aromatic side-chains seems to strongly influence the dissociation behavior in the present study, as noted in **Figure 6.1A** and **Figure 6.2**, and as also noted previously for UVPD at 266 nm.[35]

In **Figure 6.1B**, the photodissociation efficiency is shown for a series of peptides containing the same number and types of aromatic amino acids (one phenylalanine and one tyrosine), while the lengths and charge states of the peptides are varied. Each peptide from the shortest (DRVYIHPF, 8 residues) to the longest (DAEFRHDSGYQVHHQK, 16 residues) yielded nearly identical photodissociation efficiencies, ranging from 69 to 85% even when the total number of amino acids doubled. Also, as the charge state increased, the photodissociation efficiencies increased, presumably due to some combination of greater proton mobility and greater coulombic repulsion, an effect also observed previously for IR photodissociation.[36] The results in **Figure 6.1** reinforce that the number of aromatic side-chains and the precursor charge state influence photodissociation efficiency more so than the number of amide chromophores.

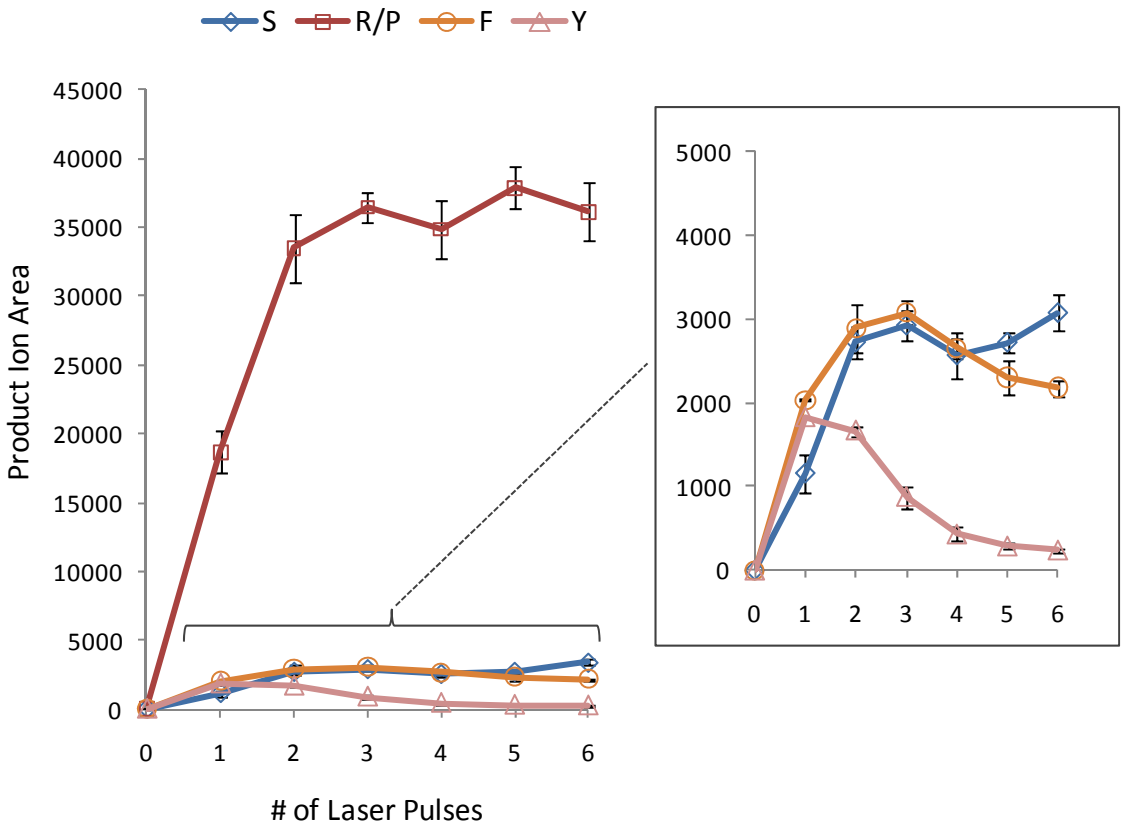

Figure 6.2 Immonium ion product area versus the number of laser pulses (5 ns, 8 mJ per pulse) for YRPPGFSPFR (2+) at a *q*-value of 0.045. The areas of arginine (R) and proline (P) immonium ions are summed together since their nominal *m/z* values are isobaric. The inset is zoomed in on the product ion areas of serine (S), phenylalanine (F), and tyrosine (Y). G immonium ions are not typically observed and thus are not monitored. The R and P immonium ions have the same *m/z* values and thus are monitored collectively.

6.4.2 Impact of Number of Laser Pulses and RF Trapping Voltage on UVPD

The influence of the number of laser pulses and the rf trapping voltage (as reflected by the *q*-value) on peptide dissociation was examined for the doubly- and triply-charged state of YRPPGFSPFR (**Figure 6.3**). The percentages of sequence, immonium, and precursor ions (based on ion peak areas) were tracked as a function of the varied parameter (i.e., number of pulses or *q*-value). Upon increasing the number of laser pulses

while maintaining a constant *q*-value of 0.05, the abundances of sequence ions increased up until two laser pulses for the doubly-charged species (**Figure 6.3A**) and up until between one and two laser pulses for the triply-charged species (**Figure 6.3B**), after which the abundances of these ions decreased. As the abundances of the precursor and sequence ions decreased with increasing laser pulses, immonium ions dominated the spectra especially after the fourth pulse for the doubly-charged species and the third pulse for the triply-charged species; this trend is attributed to their formation and survival upon UVPD of the precursor ions as well as ongoing secondary dissociation of primary sequence ions into immonium ions.

Upon increasing the *q*-value while keeping the number of laser pulses constant (using a single 8 mJ, 5 ns pulse), higher *q*-values (especially increasing from 0.05 to 0.1) resulted in a greater extent of precursor dissociation and more sequence information without significant loss of low mass diagnostic ions upon activation of both the doubly- and triply-charged precursors (**Figure 6.3C** and **Figure 6.3D**). However, at *q*-values above approximately 0.15, many low mass sequence and immonium ions were lost due to limitations associated with the low mass cutoff (LMCO) inherent to ion traps. Even for low *m/z* precursors (e.g., peptides in high charge states) where there is no significant LMCO, using higher *q*-values (0.1 versus 0.05) resulted in increased sequence information as seen for triply-charged ASHLGLAR in **Figure 6.4**. This increase in sequence information at higher *q*-values likely arises from the reduction in the size of the ion cloud and thus its better overlap with the laser beam in the ion trap, which has been studied in depth previously for IR photodissociation.[37]

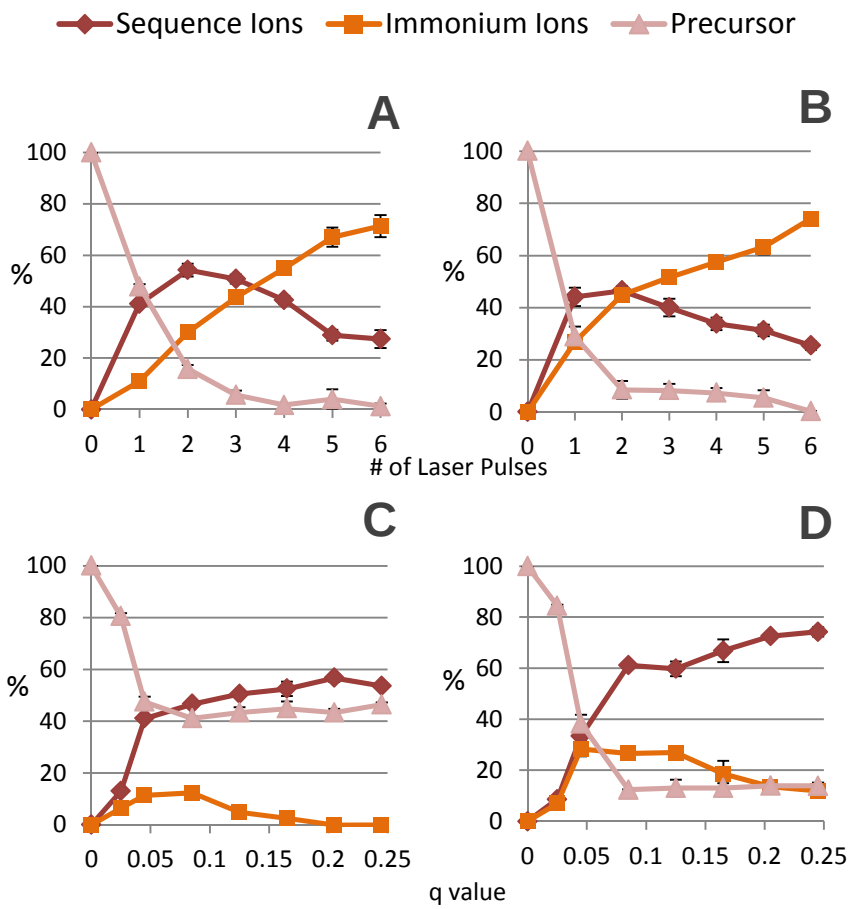

Figure 6.3 Percent abundances (as measured by peak area) of diagnostic and precursor ions versus number of laser pulses for (a) [YRPPGFSPFR + 2H]$^{2+}$ (b) [YRPPGFSPFR + 3H]$^{3+}$. Percent abundances (as measured by peak area) of diagnostic and precursor ions versus *q*-value for (c) [YRPPGFSPFR+ 2H]$^{2+}$ (d) [YRPPGFSPFR+ 3H]$^{3+}$. The *q*-value was held at 0.045 for pulse variable experiments. For variable *q*-value experiments, the activation was set to a single 5 ns, 8 mJ pulse.

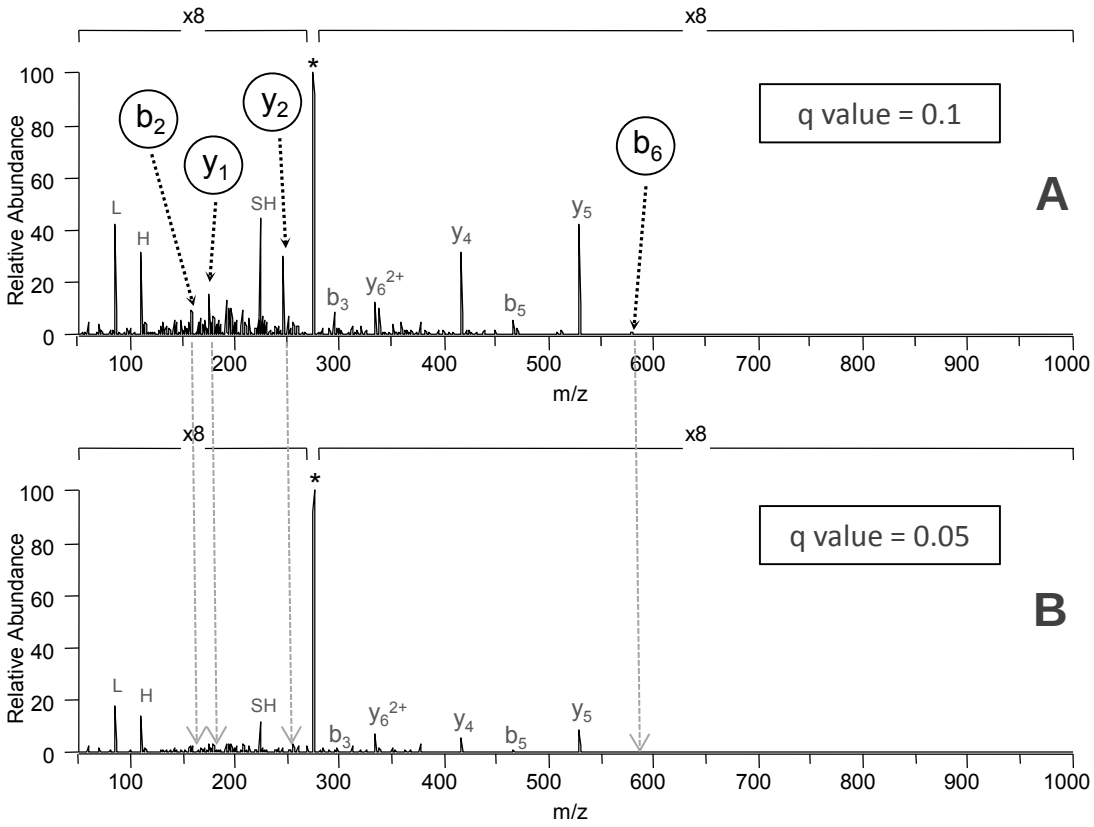

Figure 6.4 UVPD spectra of the model peptide ASHLGLAR (3+) using an activation of a single 5 ns, 8 mJ pulse at 193 nm and a *q*-value of (a) 0.1 and (b) 0.05. Circled sequence ions are unique to the spectrum collected at a *q*-value of 0.1. All other peaks (product ions) were observed in both spectra (*q*-value = 0.1 and *q*-value = 0.05). The precursor is denoted by *.

To afford the most informative tandem mass spectral information (with respect to sequence and immonium ions) in the most high-throughput manner, a single 5 ns (8 mJ) laser pulse and a *q*-value of 0.1 was used as the universal set of dissociation parameters for the remainder of the LC-MS/UVPD study. Using these activation parameters, full sequence coverage from both N-terminal and C-terminal ions was achieved from a rich array of diagnostic ions for two tryptic-like peptides, YRPPGFSPFR (3+) and ASHLGLAR (2+), each which have significantly different dissociation efficiencies as

previously described (see **Figure 6.5** for spectra). Furthermore, the presence of the side-chain loss sequence ions, w_3 and w_5, from ASHLGLAR (**Figure 6.5B**) allows the differentiation of the isobaric leucine and isoleucine residues; however, this ion type has yet to be incorporated into current *in silico* algorithms.

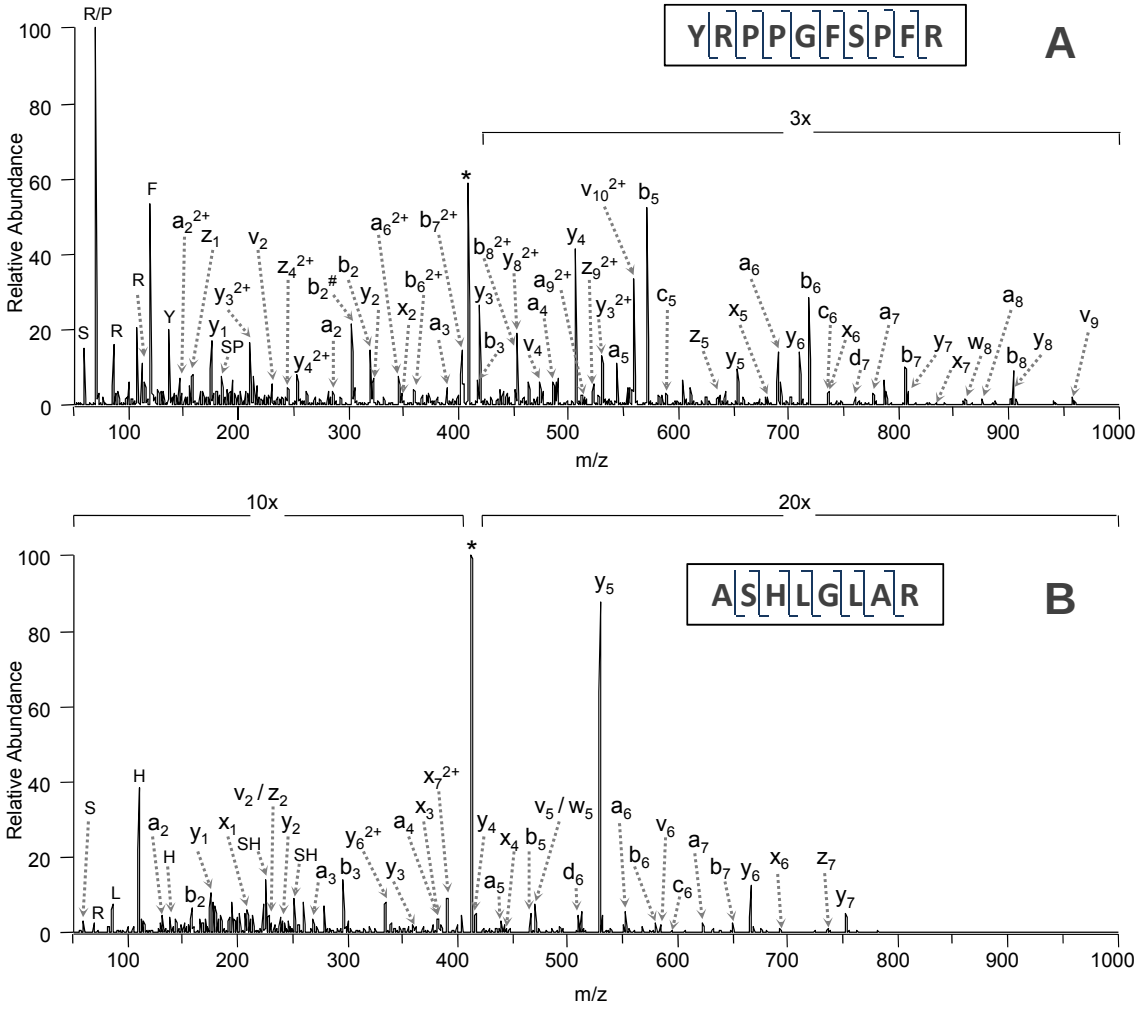

Figure 6.5 UVPD spectra of the model peptides (**a**) YRPPGFSPFR, 3+ and (**b**) ASHLGLAR, 2+. A *q*-value of 0.1 and an activation of one 5 ns, 8 mJ pulse at 193 nm were used for each spectrum. The precursor is denoted by *.

6.4.3 LC-MS/UVPD with Background Subtraction for *In Silico* Interpretation

Upon introducing 193 nm photons into the linear ion trap, significant ion abundances are seen from photoionization products. Similar products have been observed previously using 157 nm UVPD and are thought to originate from photoionization of background organic species in the vacuum system.[30] This photoionization phenomenon is illustrated in **Figure 6.6**. With no sample solution being infused or analyte ions injected into the ion trap, the region of the mass spectrum from *m/z* 496 to 504 was isolated to empty the trap of any residual ions (**Figure 6.6A**). Upon subsequent irradiation of the trapping volume using one 5 ns (8 mJ) laser pulse, significant background photoionization products are observed around ~*m/z* 200 (**Figure 6.6B**). These products are problematic because they overlap with potential low *m/z* diagnostic ions, impeding *in silico* spectral interpretations and reducing sensitivity. Therefore, a procedure for photoionization background subtraction was implemented and applied to subsequent LC-MS/UVPD runs in conjunction with standard *in silico* database searching (via SEQUEST). The background spectrum must be representative of the types and abundances of background ions formed during UVPD spectra, thus requiring acquisition using a similar scan program (i.e., similar trapping conditions and time segments). To achieve this goal, background spectra were collected using various *m/z* isolation windows (e.g., 250 $\pm$ 4, 500 $\pm$ 4, 750 $\pm$ 4, and 1000 $\pm$ 4) in an effort to find a universal background spectrum that could be applied to entire LC-MS/MS runs and also avoid loss of ions due to the LMCO. Photoionization abundances differed between the various *m/z* isolation windows used as seen in **Figure 6.7** due to variations in trapping efficiencies.

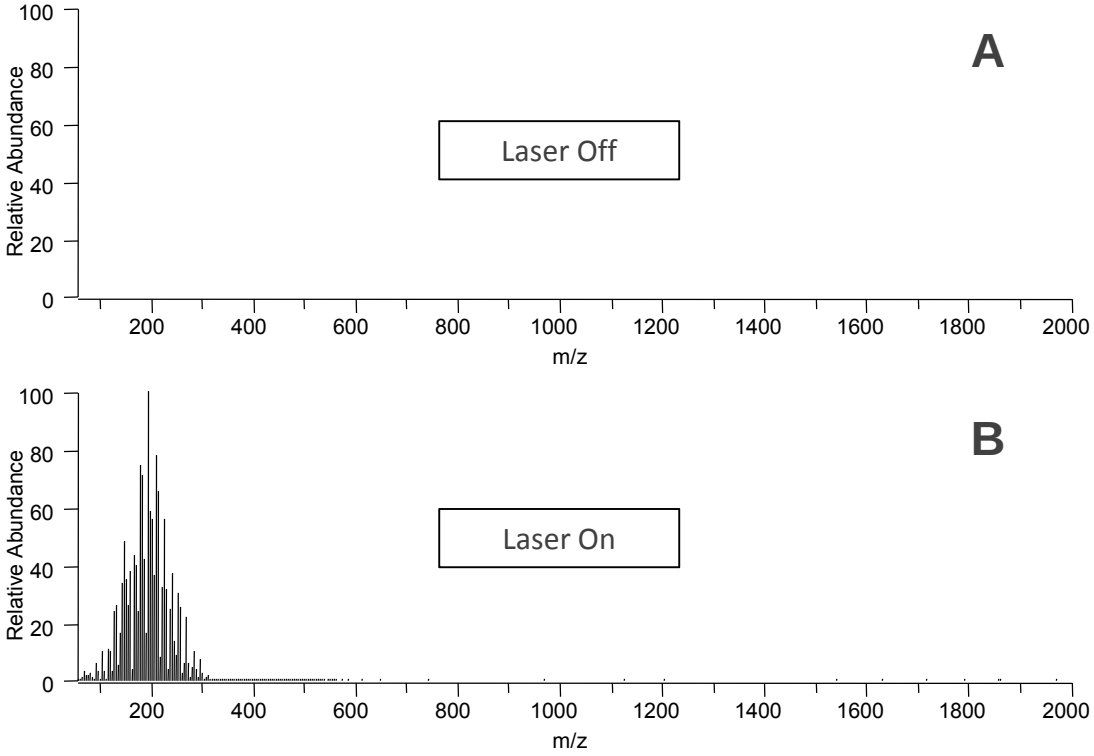

Figure 6.6 (a) Isolation of *m/z* 500 ± 4 with no sample solution infusion and the laser off and (b) isolation of *m/z* 500 ± 4 with no solution infusion and the laser on using an activation of one 5 ns (8 mJ) pulse and *q*-value of 0.1.

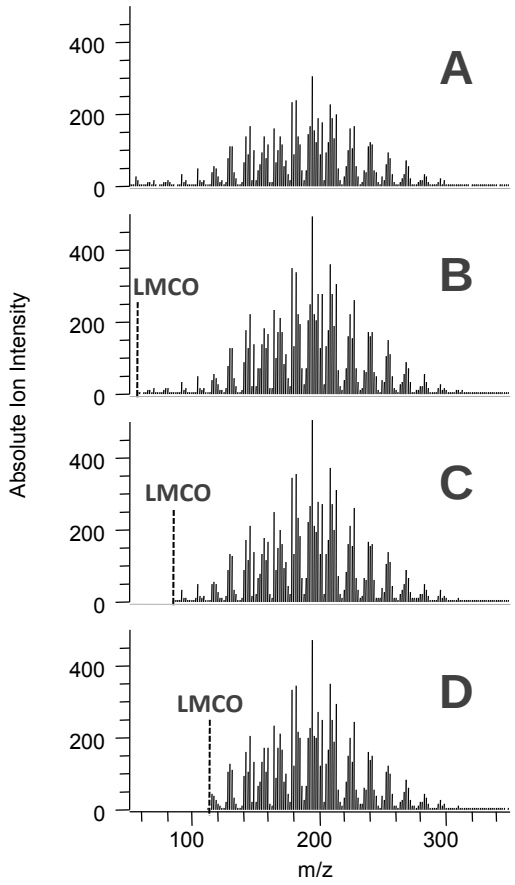

Figure 6.7 Background photoionization spectra using an isolation of (a) *m/z* 250 $\pm$ 4, (b) *m/z* 500 $\pm$ 4, (c) *m/z* 750 $\pm$ 4, and (d) *m/z* 1000 $\pm$ 4. Each spectrum was collected with no sample solution infusion, and UVPD activation of one 5 ns (8 mJ) pulse and the *q*-value was set to 0.1.

The effectiveness of the background subtraction on the UVPD spectra can be seen in **Figure 6.8** for the BSA tryptic peptide LGEYGFQNALIVR (2+) using an isolation window of *m/z* 500 $\pm$ 4. **Table 6.1** shows the SEQUEST results obtained for the tryptic digest of BSA analyzed by LC-MS/UVPD both without background subtraction and with background subtraction using background spectra acquired using the four different *m/z* isolation windows as described above (statistics were based on three runs). For each SEQUEST score (Xcorr and probability), a higher confidence is associated with a higher

score (this scoring system will be described in more detail later in this study). The use of background subtraction applied to entire LC-MS/UVPD runs significantly enhanced SEQUEST identifications and scores (XCorr and probability) for all four background subtraction options as compared to using no background subtraction procedure. Background subtraction of the photoionization spectrum acquired by isolating *m/z* 500, 750, and 1000 all outperformed the *m/z* 250 isolation window. The background spectra acquired using the *m/z* 250 window yielded lower photoionization product abundances compared to the other three isolation windows, thus resulting in a less effective subtraction of the background ions from the UVPD spectra. From these results, background subtraction using photoionization spectra at isolation windows of *m/z* 500 $\pm$ 4, 750 $\pm$ 4, or 1000 $\pm$ 4 were applied individually to all LC-MS/UVPD spectra for the remainder of the study to improve algorithm confidence and peptide identification. The best *in silico* searching results were then used for CID comparisons in subsequent sections.

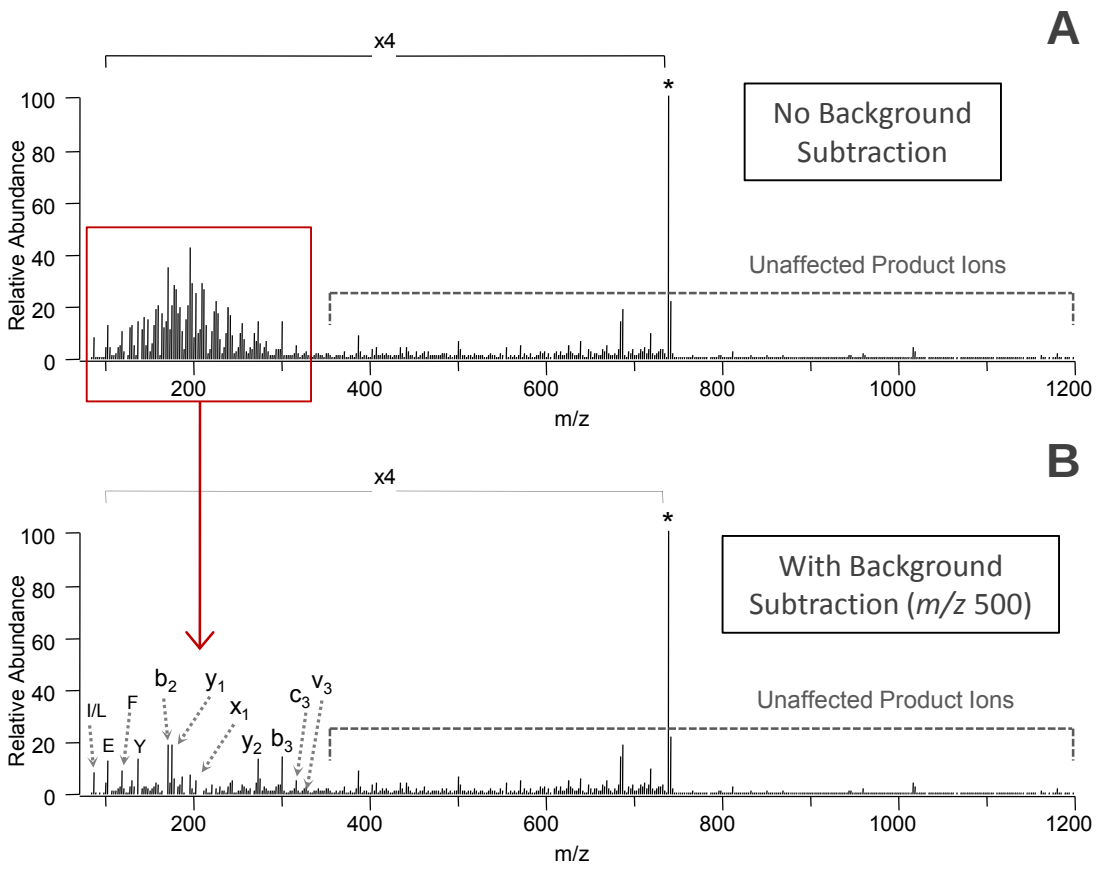

Figure 6.8 LC-MS/UVPD spectrum of the BSA peptide LGEYGFQNALIVR (2+) with (a) no background subtraction and (b) with background subtraction. A *q*-value of 0.1 and a photoactivation of one 5 ns, 8 mJ pulse were used for each spectrum. Background subtraction was achieved by subtracting the photoionization background spectrum (photoactivation after isolation of *m/z* 500 $\pm$ 4) from the entire LC-MS/UVPD run. The precursor is denoted by *, and each spectrum is the average of three microscans.

	No background Subtraction	*m/z* 250	*m/z* 500	*m/z* 750	*m/z* 1000
Sequence Coverage	66 ± 4%	70 ± 3%	74 ± 3%	75 ± 4%	76 ± 5%
Max Sequence Coverage	69%	74%	77%	78%	81%
Protein Score	409 ± 19	555 ± 33	683 ± 20	709 ± 20	677 ± 25
Max Protein Score	425.26	583.26	699.08	729.49	705.51
Number of Peptides	111 ± 7	141 ± 9	171 ± 8	195 ± 21	198 ± 17
Unique Peptides	41 ± 3	43 ± 1	47 ± 4	48 ± 3	49 ± 5
Max Peptide Xcorr Score	7.48	7.73	7.44	7.21	7.42
∑ Peptide Xcorr Score	371 ± 20	471 ± 22	561 ± 22	623 ± 55	632 ± 38
Max Peptide Prob. Score	128.26	140.85	300.00	300.00	130.40
∑ Peptide Prob. Score	3055 ± 85	3977 ± 196	4807 ± 136	5248 ± 248	5150 ± 263

Table 6.1 SEQUEST analysis of LC-MS/UVPD of a BSA tryptic digest without background subtraction, and with background subtraction using background spectra from various isolation *m/z*. Each isolation (*m/z* 250, *m/z* 500, *m/z* 750, and *m/z* 1000) has a width of ± 4 u. For each LC-MS/UVPD run, an activation of one 5 ns (8 mJ) pulse and *q*-value of 0.1 was used. Background spectra were collected with no ESI infusion at the beginning of the first of three LC-MS/UVPD runs. For each type of SEQUEST score, a higher confidence is associated with a higher score.

6.4.4 LC-MS/MS: Comparing UVPD and CID

A data-dependent LC-MS/MS method was developed that enabled alternating UVPD and CID scans of each peptide ion and allowing the most direct comparison of the attributes of the resulting tandem mass spectra upon application of the search algorithms. SEQUEST comparisons of the resulting LC-MS/MS data acquired for a tryptic digest of BSA are summarized in **Table 6.2**. A major reason SEQUEST was selected as the search algorithm over other algorithms such as MASCOT and OMSSA is that all major product

ion types (e.g., *a*, *b*, *c*, *x*, *y*, *z*) can be searched simultaneously in each spectrum. This capability directly benefits peptide identification and scoring for UVPD since Xcorr scores (the major peptide scoring system of SEQUEST) are a measure of how well a theoretical spectrum matches an experimental spectrum, and UVPD yields richer spectra with more *a*, *b*, *c*, *x*, *y*, and *z* ions compared to the array of ions obtained upon CID (typically *b* and *y*). In general, the UVPD spectra lead to significantly higher peptide Xcorr scores relative to those arising from the analogous CID spectra, often resulting in a higher number of peptide identifications (comparing the first column with the third column of **Table 6.2**). Examples of spectra illustrating this outcome are shown in **Figure 6.9** for the BSA peptide RPCFSALTPDETYVPK (3+) identified via LC-MS/MS. UVPD generated significantly more sequence ions and greater coverage than did CID, the latter which suffered both from the loss of diagnostic ions from the ion trap due to the low mass cutoff and missed backbone cleavages. The resulting Xcorr scores for RPCFSALTPDETYVPK (3+) from UVPD and CID were 4.90 and 3.58, respectively.

	UVPD (*a*,*b*,*c*,*x*,*y*,*z*)	UVPD (*b*,*y*)	CID (*b*,*y*)	CID (*a*,*b*,*c*,*x*,*y*,*z*)
Number of Peptides	100 ± 3	73 ± 8	80 ± 5	83 ± 3
Number of Peptides (1.1 probability filter)	80 ± 4	68 ± 7	77 ± 3	79 ± 1
Unique Peptides	37 ± 1	30 ± 1	35 ± 2	33 ± 3
Unique Peptides (1.1 probability filter)	33 ± 1	29 ± 1	34 ± 2	32 ± 3
Max Peptide Xcorr Score	7.19	5.16	5.17	5.61
∑ Peptide Xcorr Score	309 ± 16	193 ± 18	231 ± 2	231 ± 11
Max Peptide Prob. Score	300.00	126.79	129.22	126.79
∑ Peptide Prob. Score	2844 ± 187	2472 ± 174	2851 ± 59	3067 ± 117

Table 6.2 SEQUEST results of LC-MS/UVPD/CID of a BSA tryptic digest based on triplicate runs. UVPD and CID were both searched using *a*,*b*,*c*,*x*,*y*,*z* and *b*,*y* product ions as indicated in the table. An activation of one 5 ns (8 mJ) pulse, and *q*-value of 0.1 was used for UVPD. For each SEQUEST score, a higher confidence is associated with a higher score.

To investigate whether this increase in scoring was a direct result of increased spectral information upon UVPD or rather an artifact of searching using a greater array of sequence ions (i.e., searching with *a*, *b*, *c*, *x*, *y*, *z* for UVPD versus *b*, *y* for CID), both the UVPD and CID spectra were searched first using *a*, *b*, *c*, *x*, *y*, *z* and then again using only *b*, *y* product ions as seen in **Table 6.2**. For the UVPD data, searching with more ions dramatically increased peptide identifications and scoring as compared to searching with only *b* and *y* ions (comparing the first and second columns of **Table 6.2**). There was no significant difference within error (statistics based on three runs), however, when searching with *a*, *b*, *c*, *x*, *y*, *z* versus just *b*, *y* for the CID data (comparing the third and

fourth columns of **Table 6.2**). Therefore, the increase in scoring and number of peptide identifications obtained from the UVPD data is likely the result of more informative MS/MS spectra that contain a greater array of sequence ions than CID.

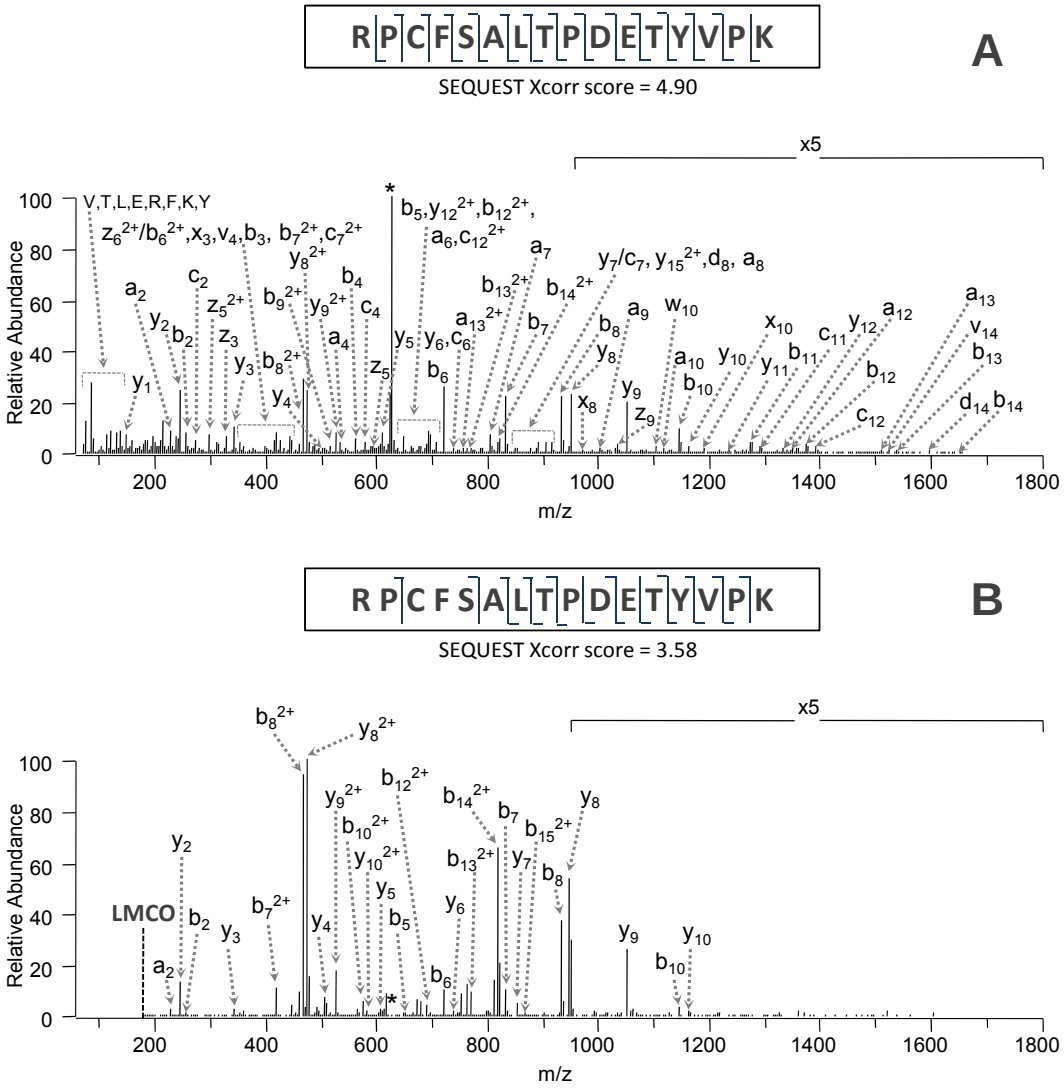

Figure 6.9 (a) UVPD spectrum (193 nm) of the BSA peptide RPCFSALTPDETYVPK (3+), *q*-value of 0.1 and an activation of one 5 ns, 8 mJ pulse. (b) CID spectrum of BSA peptide RPCFSALTPDETYVPK (3+), *q*-value of 0.25, 30 ms activation, 35% NCE. Each spectrum is the average of three microscans. The precursor is denoted by *. The UVPD and CID spectra were collected from back-to-back LC-MS/UVPD/CID data-dependent events. The cysteine in the peptide has a carbamidomethyl modification.

One drawback with database searching using many sequence ions is the potential for an increase in false positives even after filtering at a 1% false discovery rate (FDR). The majority of these false positives that passed the 1% FDR yielded low probability scores of 1.00, indicating a high probability that the peptide was identified by chance, and thus could be eliminated by additional filtering based on a probability score threshold. While this procedure is effective, it also eliminates many of the true positives identified by UVPD. Many of these true positives with low probability scores were a result of singly charged peptides producing many abundant side-chain loss sequence ions (e.g., *v*, *d*, *w*, *m*). For example, UVPD of the singly-charged, BSA peptide YLYEIAR (**Figure 6.10**) yielded a good Xcorr score of 3.04, but a poor probability score of 1.00. Manual interpretation of this spectrum showed that seven of the eleven most abundant sequence ions were *v*, *w*, and *m* side-chain loss ions with many even lower abundance side-chain loss ions also present in the spectrum. While side-chain loss ions are very useful for sequencing peptides and differentiating between leucine and isoleucine residues, they impede current *in silico* database searching algorithms, such as SEQUEST, which lack the ability to identify these ions. As observed herein and previously using UVPD at 157 nm,[30] the formation of side-chain loss ions was most notable for singly-charged peptides which are less often observed in LC-ESI-MS due to the use of acidic and aqueous eluents that usually produce peptides in charge states greater than or equal to 2+. Regardless, even after further filtering of true positive peptide hits based from low probability scores, UVPD yielded comparable and often better SEQUEST results as compared to CID.

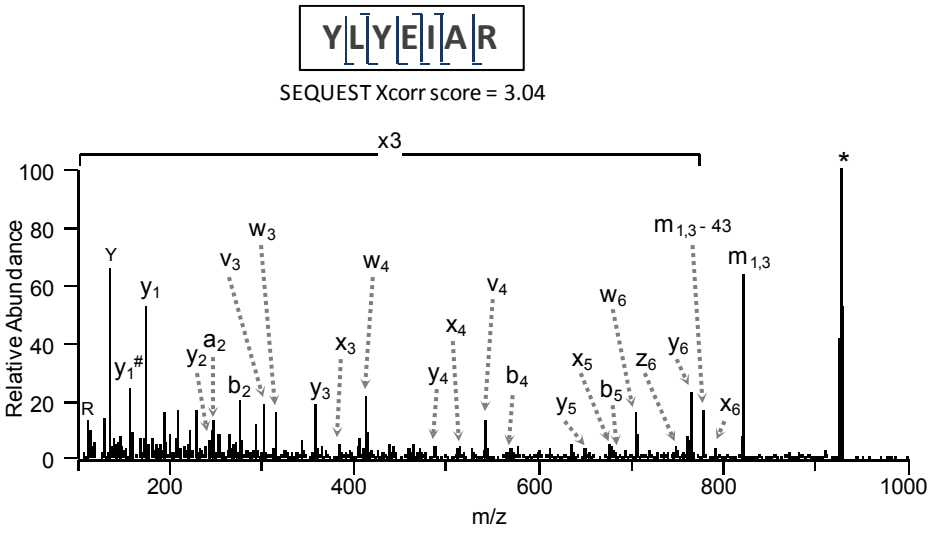

Figure 6.10 UVPD spectrum of the BSA peptide YLYEIAR (1+), *q*-value of 0.1 and an activation of one 5 ns, 8 mJ pulse. The precursor is denoted by *, and the neutral loss of ammonia is denoted by #. The ion labeled as "$m_{1,3}$" represents the side-chain loss of tyrosine radicals from the precursor, and "$m_{1,3}$ - 43" represents sequential loss of CH_3CO radicals from "$m_{1,3}$".

To investigate the performance of the high-throughput-UVPD method for complex, biologically relevant samples, data-dependent LC-MS/UVPD/CID analysis was applied to a lysed human HT-1080 cytosolic fibrosarcoma cell sample. Examples of the UVPD and CID mass spectra obtained for one peptide from this facet of the study are shown in **Figure 6.11**. **Figure 6.12** illustrates the SEQUEST results for the top twenty-five protein hits; the protein names and further SEQUEST scoring information are summarized in **Table 6.3**. All UVPD and CID spectra for which peptides were successfully identified through SEQUEST searching were manually separated and verified for each protein. As seen in **Figure 6.12A**, both UVPD and CID yielded comparable results with respect to peptide identifications with UVPD often affording several additional peptide identifications. The sums of the peptide Xcorr scores for each protein were significantly better for UVPD as compared to CID (**Figure 6.12B**). This

increase in scoring likely arises from the more informative spectra produced by UVPD versus CID as noted previously in this report for the BSA tryptic digest. For example, the UVPD mass spectrum of the doubly charged, HT-1080 peptide DLYANTVLSGGTTMYPGIADR (**Figure 6.11A**) shows many *a*, *b*, *c*, *x*, *y*, *z* product ions and yields full sequence coverage from both series of N-terminal and C-terminal ions. CID of the same peptide, however, yielded significantly less spectral information and incomplete peptide sequence coverage (**Figure 6.11B**). Thus, the Xcorr scores were 6.78 and 4.44 for UVPD and CID, respectively. No significant trend with respect to probability scores was observed for UVPD versus CID (**Figure 6.12C**); however, CID did outperform UVPD with respect to the summation of these scores for several proteins (e.g, Protein # 4, 5, 6, and 14) due to the presence of a few peptides with very high, max probability scores as seen in **Table 6.3**. The total protein identifications were comparable between the two methods for the back-to-back LC-MS/UVPD/CID analysis. After peptide filtering at a 1% FDR followed by further filtering of peptides with probability scores less than 1.00, CID spectra identified 191 proteins and UVPD spectra identified 202 proteins. Overall, the faster UVPD method (using a single 5 ns, 8 mJ pulse) performed comparably or often better than the slower CID approach even for complex biological samples based on this high-throughput, shotgun-style proteomic strategy.

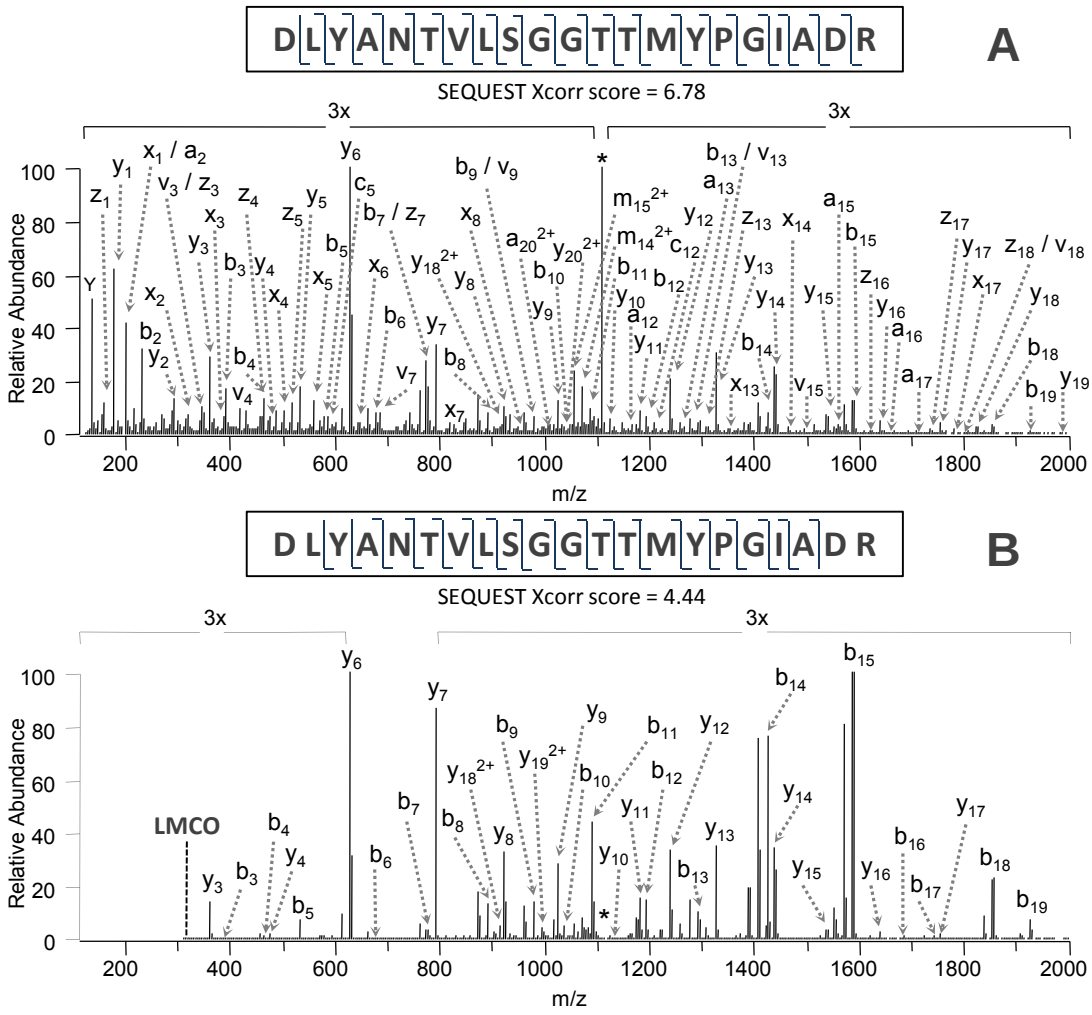

Figure 6.11 (a) UVPD spectrum (193 nm) of the beta actin peptide DLYANTVLSGGTTMYPGIADR (2+) from a human HT-1080 cytosolic lysed cell sample, *q*-value of 0.1 and an activation of one 5 ns, 8 mJ pulse. Low abundant w_4, w_6, and w_{14} side-chain loss ions were observed, but were not labeled on the spectrum. (b) CID spectrum of the beta actin peptide DLYANTVLSGGTTMYPGIADR (2+) from a human HT-1080 cytosolic cell sample, *q*-value of 0.25, 30 ms activation, 35% NCE. Each spectrum is the average of three microscans. The precursor is denoted by *. The UVPD and CID spectra were collected from back-to-back LC-MS/MS data-dependent events.

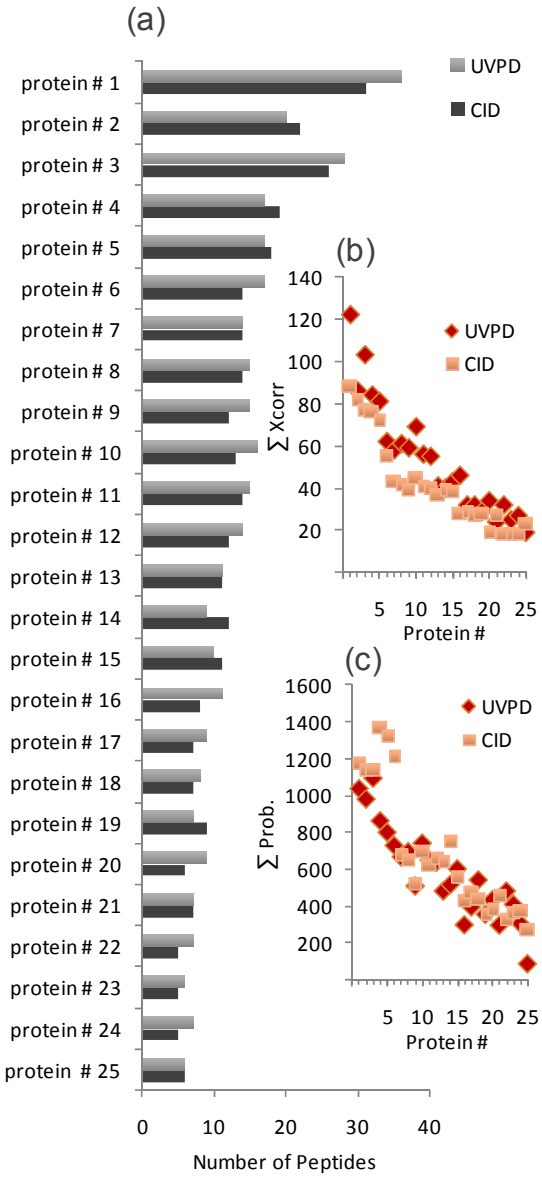

Figure 6.12 SEQUEST analysis of LC-MS/UVPD/CID of human HT-1080 cytosolic lysed cells: (**a**) number of peptides identified for UVPD versus CID for each protein (**b**) sum of peptide Xcorr scores for each protein for UVPD and CID (**c**) sum of peptide probability scores for each protein for UVPD and CID. The top twenty-five protein hits were used for manual separation and comparison of UVPD and CID data. The protein names and further scoring information are shown in Table 6.3.

	# Peptides	# Unique Peptides	Max peptide Xcorr Score	∑ Peptide Xcorr Score	Max Peptide Prob. Score	∑ Peptide Prob. Score
Protein # 1 = vimentin – gi62414289:						
UVPD	36 (30)	27 (25)	6.27	122	137.32	1037
CID	31 (30)	26 (25)	5.40	88	110.94	1174
Protein # 2 = enolase 1 – gi4503571:						
UVPD	20 (18)	18 (16)	6.64	86	128.85	981
CID	22 (20)	17 (15)	6.87	82	117.52	1139
Protein # 3 = beta actin – gi4501885:						
UVPD	28 (25)	18 (15)	6.78	103	125.04	1094
CID	26 (23)	15 (14)	5.72	77	112.65	1138
Protein # 4 = tubulin, beta polypeptide – gi29788785:						
UVPD	17 (15)	14 (12)	9.89	84	124.43	863
CID	19 (19)	15 (15)	8.15	76	300.00	1371
Protein # 5 = tubulin, beta 2 – gi5174735:						
UVPD	17 (14)	14 (11)	9.89	81	124.43	801
CID	18 (18)	14 (14)	8.15	72	300.00	1325
Protein # 6 = glyceraldehydes-3-phosphate dehydrogenase – gi7669492:						
UVPD	17 (16)	11 (10)	7.35	62	126.07	730
CID	14 (14)	10 (10)	5.97	55	300.00	1212
Protein # 7 = heat shock protein 90kDa alpha (cytosilic), class A member 1 isoform 2 – gi40254816:						
UVPD	14 (14	13 (13)	7.46	58	119.47	668
CID	14 (14)	12 (12)	4.19	43	91.48	674
Protein # 8 = heat shock protein 90kDa alpha (cytosilic), class A member 1 isoform 1 – gi63029937:						
UVPD	15 (14)	14 (13)	7.46	61	119.47	696
CID	14 (14)	13 (13)	4.19	42	91.48	654
Protein # 9 = annexin A2 isoform 2 – gi50845386:						
UVPD	15 (14)	14 (13)	6.84	59	103.47	510
CID	12 (12)	10 (10)	4.26	39	93.63	519
Protein # 10 = pyruvate kinase 3 isoform 1 – gi33286418:						
UVPD	16 (15)	13 (12)	7.07	69	128.64	742
CID	13 (12)	12 (11)	5.42	45	111.12	692
Protein # 11 = heat shock protein 90kDa protein 1, beta – gi20149594:						
UVPD	15 (14)	13 (13)	6.76	56	110.15	650
CID	14 (14)	13 (13)	4.32	41	101.63	626
Protein # 12 = eukaryotic translation elongation factor 2 – gi4503483:						
UVPD	14 (10)	14 (10)	7.08	55	127.83	639
CID	12 (11)	12 (11)	4.97	40	119.12	658
Protein # 13 = tubulin, beta, 4 – gi50592996:						
UVPD	11 (9)	11 (9)	5.36	41	98.45	483
CID	11 (11)	10 (10)	4.28	37	93.97	637
Protein # 14 = lactate dehydrogenase A – gi5031857:						
UVPD	9 (6)	9 (6)	7.36	40	140.25	518
CID	12 (12)	12 (12)	6.32	39	300.00	751
Protein # 15 = tubulin alpha 6 – gi14389309:						
UVPD	10 (10)	10 (10)	6.55	43	115.67	603
CID	11 (11)	11 (11)	5.65	38	102.71	560
Protein # 16 = eukaryotic translation factor elongation factor 1 alpha 1 – gi4503471:						
UVPD	11 (10)	9 (8)	7.13	46	97.79	300
CID	8 (8)	6 (6)	5.75	28	119.59	429
Protein # 17 = phosphoglycerate kinase 1 – gi4505763:						
UVPD	9 (7)	8 (7)	5.85	32	87.78	397
CID	7 (7)	7 (7)	5.62	29	110.55	474
Protein # 18 = myosin, heavy polypeptide 9, non-muscle – gi12667788:						
UVPD	8 (7)	8 (7)	5.75	32	121.89	543
CID	7 (7)	7 (7)	5.89	27	116.48	442
Protein # 19 = heat shock protein 70kDa 8 isoform 1/2 – gi5729877 / gi24234686:						
UVPD	7 (6)	7 (6)	6.57	29	132.10	357
CID	9 (9)	8 (8)	5.78	28	103.39	353
Protein # 20 = actinin, alpha 4 – gi12025678:						
UVPD	9 (7)	9 (7)	5.02	34	92.21	435
CID	6 (6)	6 (6)	4.08	19	85.42	383
Protein # 21 = aldolase A – gi34577112:						
UVPD	7 (7)	7 (7)	5.84	24	98.73	298
CID	7 (7)	7 (7)	5.53	27	122.45	453

Protein # 22 = cofilin 1 (non-muscle) – gi5031635:						
UVPD	7 (5)	6 (4)	9.07	32	300.00	483
CID	5 (5)	4 (4)	6.65	18	115.02	330
Protein # 23 = actinin alpha 1 – gi4501891:						
UVPD	6 (6)	6 (6)	4.87	25	105.62	411
CID	5 (5)	5 (5)	4.14	18	95.53	367
Protein # 24 = triosephosphate isomerase 1 – gi4507645:						
UVPD	7 (6)	7 (6)	7.12	27	88.44	307
CID	5 (5)	5 (5)	4.81	18	129.37	375
Protein # 25 = talin 1 – gi16753233:						
UVPD	6 (2)	6 (2)	5.24	19	81.96	89
CID	6 (6)	6 (6)	5.53	23	105.58	269

Table 6.3 UVPD and CID comparison of the number of peptides identified, sum and max peptide Xcorr score, and sum and mass peptide probability score from SEQUEST for the top twenty-five protein hits of a human HT-1080 cytosolic cell lysate sample. The UVPD and CID spectra were collected from back-to-back LC-MS/UVPD/CID data-dependant events. Numbers in parenthesis represent the number of peptide identifications after filtering out peptides with a probability score less than 1.1. For each SEQUEST score, a higher confidence is associated with a higher score. NCBI sequence identifiers ("gi" numbers) are listed next to each protein.

6.5 Conclusions

The results in the present study illustrate the utility of using ultrafast UVPD for high-throughput proteomic workflows. UVPD at 193 nm results in ample production of a, b, c, x, y, and z sequence ions, in addition to immonium ions and v and w side-chain loss ions. Photodissociation efficiencies range from 50-98% for doubly-charged peptides using a single laser pulse. Peptides containing one or more amino acids with aromatic side-chains exhibit higher dissociation efficiencies than peptides without aromatic groups, and the length of the peptide has relatively little impact on UVPD efficiency. A background subtraction procedure to account for the formation of rather high abundances of ions due to photoionization of background organic species during the UVPD period was implemented and significantly improved subsequent SEQUEST scoring and peptide identifications for complex mixtures analyzed by LC-UVPD-MS. Comparable and often improved *in silico* searching results were achieved for biologically relevant samples, including a human HT-1080 cytosolic fibrosarcoma cell sample, using ultrafast UVPD in direct comparison to the gold standard method of CID.

6.6 References

(1) de Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Froehlich, F.; Walther, T. C.; Mann, M. *Nature* **2008**, *455*, 1251-1254.

(2) Eng, J. K.; McCormack, A. L.; Yates, J. R., III *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 976-989.

(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

(4) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *Journal of Proteome Research* **2004**, *3*, 958-964.

(5) Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466-1467.

(6) Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. *Molecular and Cellular Proteomics* **2007**, *6*, 1599-1608.

(7) Hardman, M.; Makarov, A. A. *Analytical Chemistry* **2003**, *75*, 1699-1705.

(8) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. *Journal of Mass Spectrometry* **2005**, *40*, 430-443.

(9) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Analytical Chemistry* **2006**, *78*, 2113-2120.

(10) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. *Molecular and Cellular Proteomics* **2009**, *8*, 2759-2769.

(11) Pekar Second, T.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. *Analytical Chemistry* **2009**, *81*, 7757-7765.

(12) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.

(13) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 2199-2204.

(14) Swaney, D. L.; Wenger, C. D.; Thomson, J. A.; Coon, J. J. *Proceedings of the National Academy of Sciences of the United States of America, Early Edition* **2009**, 1-6, 6 pp.

(15) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. *Analytical Chemistry* **2008**, *80*, 4825-4835.

(16) Swaney, D. L.; Wenger, C. D.; Coon, J. J. *Journal of Proteome Research* **2010**, *9*, 1323-1329.

(17) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nature Methods* **2007**, *4*, 709-712.

(18) Brodbelt, J. S.; Wilson, J. J. *Mass Spectrometry Reviews* **2009**, *28*, 390-424.

(19) Reilly, J. P. *Mass Spectrometry Reviews* **2009**, *28*, 425-447.
(20) Ly, T.; Julian, R. R. *Angewandte Chemie, International Edition* **2009**, *48*, 7130-7137.
(21) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. *Analytical Chemistry* **2009**, *81*, 8109-8118.
(22) Bowers, W. D.; Delbert, S. S.; Hunter, R. L.; McIver, R. T., Jr. *Journal of the American Chemical Society* **1984**, *106*, 7288-7289.
(23) Hunt, D. F.; Shabanowitz, J.; Yates, J. R., III *Journal of the Chemical Society, Chemical Communications* **1987**, 548-550.
(24) Hunt, D. F.; Shabanowitz, J.; Yates, J. R., III; Griffin, P. R.; Zhu, N. Z. *Analytica Chimica Acta* **1989**, *225*, 1-10.
(25) Lebrilla, C. B.; Wang, D. T. S.; Mizoguchi, T. J.; McIver, R. T., Jr. *Journal of the American Chemical Society* **1989**, *111*, 8593-8598.
(26) Zhang, L.; Reilly, J. P. *Analytical Chemistry* **2010**, *82*, 898-908.
(27) Kim, T.-Y.; Reilly, J. P. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 2334-2341.
(28) Shin, Y. S.; Moon, J. H.; Kim, M. S. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 53-59.
(29) Thompson, M. S.; Cui, W.; Reilly, J. P. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1439-1452.
(30) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 1657-1665.
(31) Morgan, J. W.; Russell, D. H. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 721-729.
(32) Zhang, L.; Reilly, J. P. *Journal of Proteome Research* **2010**, *9*, 3025-3034.
(33) Kalcic, C. L.; Gunaratne, T. C.; Jones, A. D.; Dantus, M.; Reid, G. E. *Journal of the American Chemical Society* **2009**, *131*, 940-942.
(34) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80*, 4807-4819.
(35) Oh, J. Y.; Moon, J. H.; Kim, M. S. *Journal of Mass Spectrometry* **2005**, *40*, 899-907.
(36) Madsen, J. A.; Brodbelt, J. S. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 349-358.
(37) Remes, P. M.; Glish, G. L. *Journal of Physical Chemistry A* **2009**, *113*, 3447-3454.

Chapter 7

193 nm Photodissociation for Acidic Proteome Characterization

7.1 Overview

193 nm ultraviolet photodissociation (UVPD) was implemented to sequence singly and multiply charged peptide anions. Upon dissociation by this method, *a*-/*x*-type, followed by *d* and *w* side-chain loss ions, were the most prolific and abundant sequence ions, often yielding 100% sequence coverage. The dissociation behavior of singly and multiply charged anions was significantly different with higher charged precursors yielding more sequence ions; however, all charge states investigated (1- through 3-) produced rich diagnostic information. UVPD at 193 nm was also shown to successfully differentiate and pinpoint labile phosphorylation modifications. The sequence ions were produced with high abundances, requiring limited averaging for satisfactory spectral quality. The intact, charge-reduced radical products generated by UV photoexcitation were also subjected to collision induced dissociation (termed, activated – electron photodetachment dissociation (a-EPD)), but UVPD alone yielded more predictable and higher abundance sequence ions. With the use of a basic (pH ~11.5), piperidine-modified mobile phase, LC-MS/UVPD was implemented and resulted in the successful analysis of mitogen-activated pathway kinases (MAPKs) using ultrafast activation times (5 nanoseconds).

7.2 Introduction

The advent of new high performance tandem mass spectrometers equipped with the most versatile collision- and electron-based activation methods as well as ever-more powerful database search algorithms have catalyzed tremendous progress in the field of proteomics.[1-4] Despite these advances in instrumentation and methodologies, there are few methods that fully exploit the information available from the acidic proteome. Typical high-throughput, bottom-up workflows consist of chromatographic separation of complex mixtures of digested proteins followed by online mass spectrometry (MS) and MS^n analysis. This bottom-up approach remains the most popular strategy for protein identification, biomarker discovery, quantitative proteomics, and elucidation of post-translational modifications. To date, proteome characterization by mass spectrometry has overwhelmingly focused on analysis of peptide cations,[5] resulting in an inherent bias towards basic peptides that easily ionize under acidic mobile phase conditions and positive polarity MS settings. Given that approximately 50% of peptides/proteins are naturally acidic, coupled with the fact that many of the most important PTMs (e.g., phosphorylation, acetylation, sulfonation, etc.) significanlty decrease the isoelectric points of peptides,[6, 7] there is a compelling need for better analytical methodologies for characterization of the acidic proteome.

A principal reason for the shortage in methods for peptide anion characterization is the lack of MS/MS techniques befitting for efficient and predictable dissociation of peptide anions. Despite the growing array of new ion activation methods for dissociation of peptides, most are best suited for analysis of positively-charged peptides. CID of peptide anions, for example, yields unpredictable fragmentation behavior with spectra subjugated by neutral losses from both precursor and product ions;[8] insufficient peptide

sequence information is also the result. The two most promising new electron-based methods, electron-capture dissociation (ECD) and electron-transfer dissociation (ETD), are only applicable to positively-charged ions, not anions.[9-12] Due to the known inadequacy of CID and the lack of feasibility of ECD and ETD for peptide anion sequencing, several alternative MS^n methods have been developed recently. Electron detachment dissociation (EDD) using high energy electrons to induce backbone cleavages was developed for peptide anions.[13, 14] Another new technique, negative electron transfer dissociation (NETD), entails reactions of radical cation reagents with peptide anions to promote electron transfer from the peptide to the reagent that causes radical-directed dissociation.[15, 16] Activated-electron photodetachment dissociation (a-EPD), an MS^3 technique, uses UV irradiation to produce intact peptide radical anions, which are then collisionally activated.[17, 18] Despite these inroads in the characterization of peptide anions, these methods also suffer from several significant shortcomings. EDD and a-EPD are both low efficiency methods that require long averaging cycles and activation times that range from half a second up to multiple seconds, impeding the integration of these methods with chromatographic timescales.[13-18] In addition, the fragmentation patterns frequently yield many high abundance neutral losses from product ions which clutter the spectra,[13-16] and often low peptide sequence information.[13, 17, 18]

In this chapter, we report the use of 193 nm photons (UVPD) for peptide anion activation, which yields rich and predictable fragmentation patterns with high sequence coverage on a fast liquid chromatographic timeline. This method shows promise for a range of peptide charge states as well as for both unmodified and phosphorylated species.

7.3 Experimental

7.3.1 Materials

Human aquaporin-2 (254-267), RQpSVELHSPQSLPR, was purchased from AnaSpec (Freemont, CA), and pp60c-src (521-533), TSTEPQpYQPGENL, and DRVYIHPFHLVIH were purchased from BACHEM (King of Prussia, PA). Piperidine and buffer components, α-casein, and proteomics grade trypsin were obtained from Sigma-Aldrich (St. Louis, MO). Native Pfu DNA polymerase was purchased from Strategene (La Jolla, CA). Oligonucleotides were synthesized by Integrated DNA Technologies Inc. (Coralville, IA). Qiagen Inc. (Valencia, CA) supplied Ni-NTA agarose, Quiaprep Spin miniprep, PCR QIAquick Purification Kit, and QIAEX II Gel Extraction Kit. Tryptone and yeast extract were obtained from USB corporation (Clevelend, OH). The pET28a(+) vector was purchased from Novagen and pGEX 4T1 vector was purchased from GE Healthcare. The remaining molecular biology reagents, including DNA ladders and protein molecular mass standards, were obtained from Invitrogen Corp. (Carlsbad, CA). Ultrapure grade Tris and HEPES were purchased from Sigma (St. Louis, MO), Perkin Elmer supplied [γ-^{32}P]-ATP. P81 Ion Exchange cellulose chromatography paper was purchased from Whatman. IPTG, and DTT were obtained from US Biologicals (Swampscott, MA). Restriction enzymes, PCR00 reagents, and T4 DNA ligase were from New England Biolabs (Beverly, MA). The Mono Q HR 10/10 anion exchange column and HiPrep 26/10 desalting column were purchased from Amersham Biosciences (Piscataway, NJ). 15 mL Amicon Ultra centrifugal filters and ultrafiltration membranes were from Millipore. The *Escherichia coli* strain DH5α, used for cloning and mutagenesis, and the strains BL21 (DE3) used for recombinant protein expression, were obtained from Invitrogen. All other buffer components and chemicals were obtained from Sigma (St. Louis, MO) and Fisher scientific (Fairlawn, NJ).

7.3.2 General Methods

Protocols for all the standard molecular biology manipulations including PCR, restriction enzyme digestion, ligation and transformation, were based on methods described by the manufacturer. Plasmids were transformed into cells using a BTX Transporter Plus device. UV-visible absorbance readings were taken on a Varian Cary Model 50 spectrophotometer. A Pharmacia ÄKTA^{TM} FPLC system was used for the chromatographic purification steps. Protein purity was analyzed by Tris-glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions on 12 % gels using the Bio-Rad Mini-protean III vertical gel electrophoresis apparatus. A Techne Genius Thermal Cycler was used for the polymerase chain reaction PCR (Techne Inc., Princeton, NJ). Radioactivity measurements were performed on a Packard 1500 scintillation counter. DNA constructs were verified by sequencing at the UT core facilities using an Applied Biosystem automatic DNA sequencer.

7.3.3 Preparation of Proteins

MKK7 phosphorylated by MEKK1c

Construction of pGEX4T1-MKK7-His$_6$ – A construct encoding full-length Mus musculus mitogen-activated protein kinase kinase 7 (GenBank accession number NM_011944) with an *N*-terminal cleavable GST–tag and a *C*-terminal cleavable His$_6$-tag (with the sequence, LVPRGSHHHHHH, containing a cleavage site for Thrombin protease), was created as follows. First, PCR amplification of an MKK7 template was achieved by using the following oligonucleotides: forward (5'- CCG **GAA TTC** ATG GCG GCG TCC TCC CTG GAG CAG AAG - 3') (*EcoRI* site in bold) and reverse (5'- AT**G CGG CCG C**TC TCA GTG ATG ATG ATG ATG ATG GGA TCC ACG CGG

AAC CAG CCT GAA GAA GGG CAG ATG GTG CTG - 3') (*NotI* site in bold). The PCR product was digested with *EcoRI* and *NotI*, and the resulting digested product ligated into an *EcoRI–NotI* digested pGEX-4T1 vector. The ligation mixture was then transformed into DH5α *E. coli* cells with the appropriate antibiotic and the correct construct recovered using standard molecular biology procedures.

Expression and purification of GST-MKK7-His$_6$ – The pGEX 4T1 vector containing DNA encoding full length MKK7 was transformed into BL21 (DE3) electro-competent cells. From a single colony of freshly transformed cells, a 30 mL culture of Lauria Broth (LB) containing 50 μg/mL ampicillin was inoculated and incubated with shaking overnight at 37 ºC. The culture was diluted 100-fold into TB (Terrific Broth) media containing 50 μg/mL ampicillin, and incubated at 37 ºC with shaking. Once the OD_{600} of the culture had reached 0.6, the expression was induced by 25 μM IPTG, and shaking continued at 25 ºC for 20 hours. The cells were pelleted (8000 × g, 12 min) and the bacterial pellet immediately frozen in liquid nitrogen and stored at -80 ºC. The frozen wet cells were resuspended in 150 mL of Buffer A (20 mM Tris, pH 8.0, 0.03% Brij-30, 0.1% (*v/v*) *β*-mercaptoethanol, 5 mM imidazole, 1 mM benzamidine, 0.1 mM PMSF, and 0.1 mM TPCK) containing 0.5 M NaCl, 0.2 mg/mL lysozyme, 1 mM $MgCl_2$, and 20% glycerol. The mixture was incubated at 4 ºC for 30 minutes, then Triton-100 was added to a final concentration of 1% (*v/v*) and the incubation at 4 ºC extended for another 30 minutes. The cell lysate was sonicated (at 5 s pulses with 5 s intervals) at 4 ºC for 5 minutes and then cleared by centrifugation at 18,000 × g for 30 min. The supernatant was agitated with Ni-NTA beads for 1 h at 4 °C. After washing the beads with 150 mL of buffer A containing 10 mM imidazole and 20% glycerol, the GST-MKK7-His_6 tagged protein was eluted using 25 mL buffer A (pH 8.0) containing 200 mM imidazole and 20% glycerol. The eluted protein was collected and dialyzed into buffer S (25 mM

HEPES pH 7.5, 50 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 2 mM DTT) containing 20% glycerol. The concentration was established using the Bradford reagent. The estimated yield is around 20 mg pure *GST-MKK7-His$_6$* per liter cells.

Activation of GST-MKK7-His$_6$ – 4 μM *GST-MKK7-His$_6$*, 2 μM GST-MEKK1c (*C*-terminal 320 amino acids corresponding to the catalytic domain) were incubated for 60 min at 30 °C in the presence of 4 mM ATP in 10 mL of activation buffer B (25 mM Hepes pH 7.5, 2 mM dithiothrietol, 20 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM EGTA) and re-purified using a Ni-affinity column that purifies the active GST-MKK7-His_6 away from GST-MEKK1c. The activated GST-MKK7-His_6 was stored in buffer S containing 20% glycerol at -80 ºC until further use. At least 60% of *GST-MKK7-His$_6$* has been recovered as pure active *GST-MKK7-His$_6$*.

JNK2α2 phosphorylated by MKK4 and MKK7

Construction of pET28a(+)-JNK2α2 – 424 amino acids correspond to full-length protein from cDNA of wild type Homo sapiens mitogen-activated protein kinase 9; (GenBank accession number NM_002752) with *N*-terminal cleavable His_6-tag was created using oligonucleotides: forward (5'- CTA **GCT AGC** ATG GGC GAC AGT AAA TGT GAC AGT-3') (*NheI* site in bold) and reverse (5'- TA AT**A AGC TT**G CTA CCT GAA GAA GGG CAG ATG GTG -3') (*HindIII* site in bold). The PCR product was digested with *NheI* and *HindIII*, ligated into the *NheI HindIII* digested pET28a(+)vector, and then transformed into DH5α *E. coli* cells. The construct was verified by sequencing the DNA at the UT core facilities using an Applied Biosystem automatic DNA sequencer.

Expression and purification of tagless JNK2α2–pET28a(+)-JNK2 was electroporated into the *E.coli* strain BL21(DE3). Cells from a single colony were used to inoculate 30 mL Luria-Broth media containing 30 μg/mL kanamycin and grown

overnight at 37 °C. The culture was diluted 100-fold into Luria broth containing 30 μg/mL kanamycin and were grown at 37 °C to an $O.D_{600}$ of 0.6, the cells then were induced by 500 μM IPTG and cultured for 3-5 hours at 30 °C. The cells were pelleted (8000 × g, 12 min) then lysed in 150 mL of buffer A containing 0.5 M NaCl and 1% Triton X-100, and sonicated for 20 min at 4 °C (at 5 s pulses with 5 s intervals). The lysate was centrifuged for 30 minutes at 18000 × g and the supernatant agitated with Ni-NTA beads for 1 h at 4 °C. After washing the beads with 150 mL of buffer A containing 10 mM imidazole, the His_6-tagged proteins were eluted with 25 mL buffer A (pH 8.0) containing 200 mM imidazole. The eluted proteins were applied to a Mono Q HR 10/10 anion exchange column equilibrated in buffer C (20 mM Tris pH 8.0, (v/v) 0.03% Brij-30, (v/v) 0.1% β-mercaptoethanol, 0.1 mM EDTA and 0.1 mM EGTA). The column was developed over 15-17 column volumes with a linear gradient of 0-0.5 M NaCl. Eluted fractions of *His_6-JNK2α2* were dialyzed overnight with thrombin cleavage buffer D (20 mM Tris-HCl pH 8.4, 150 mM NaCl, and 0.1 % (v/v) β-mercapto ethanol). The dialyzed protein was incubated with thrombin (1 unit thrombin/mg protein) and 2.5 mM $CaCl_2$ for 3-5 hours at room temperature with mild agitation. Cleavage was estimated by 10 % SDS/PAGE. After cleavage, the reaction mixture (10 mL) was diluted fivefold with Buffer C and the protein was filtered and loaded on a Mono Q HR 10/10 anion exchange column equilibrated in buffer C. The column was developed over 15-17 column volumes with a linear gradient of 0-0.5 M NaCl. Eluted Fractions of *tagless JNK2α2* were collected and dialyzed into buffer S containing 10% glycerol, the concentration was established using Bradford reagent and the purity assessed by 12% SDS/PAGE. The estimated yield is around 35 mg pure *tagless JNK2α2* per liter cells.

Activation of tagless JNK2α2 by active MKK4 and active MKK7 – 2 μM tagless *JNK2α2*, 100 nM active *GST-MKK4* and 400 nM active *GST-MKK7-His_6* were incubated

for 60 min at 30 °C in presence of 4 mM ATP in 10 mL activation buffer B and re-purified using Gel filtration column (120 ml HiLoad 16/60 Superdex 200 prep grade column) equilibrated in 25 mM HEPES buffer (pH 7.5) containing 100 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA and 2 mM TCEP. The activated, tagless *JNK2α2* was stored in buffer S containing 10% glycerol at -80 ºC until further use. At least 50% of tagless *JNK2α2* has been recovered as active tagless *JNK2α2.*

ERK2 phosphorylated by MKK1

Expression, purification and activation of ERK2 – Activated His_6-tagged ERK2 (Rattus norvegicus mitogen activated protein kinase 1 with GenBank accession number NM_053842) was generated as essentially as described previously by Waas et al.[1] with minor modifications. Cleavage of the His_6 tag before activation was achieved by subjecting 10 mg of purified (His_6-tagged) inactive ERK2 to thrombin cleavage as described above for His_6-JNK2α2. The tagless ERK2 was recovered using a MonoQ column, the protein fractions were pooled and dialyzed overnight at 4 °C into buffer S containing 10% glycerol, then used for activation following the same previous protocol.

p38 MAPKα phosphorylated by MKK6

Expression and purification of tagless p38 MAPKα – A pET14B vector containing DNA sequences encoding Mus musculus mitogen-activated protein kinase 14 - p38 MAPKα (GenBank accession number NM_011951) was used to express p38 MAPKα as an *N*-terminal, His_6-tagged protein in E. coli BL21 (DE3) pLysS. The enzyme was expressed and purified according to the method of Szafranska et al.[19, 20] The His_6-tag was removed by Thrombin (Novagen) and the enzyme preparation dialyzed overnight at 4 °C into storage buffer S containing 5% glycerol and stored at -80 ºC until further use.

Expression and purification of GST-MKK6b (S207E T211E) – MKK6b (S207E T211E) was expressed as a GST fusion protein in E. coli BL21 (DE3) pLysS. The

protein was purified by Glutathione Sepharose™ High Performance (Amersham Biosciences) following exactly the same method of Szafranska et al.[19, 20] Pooled protein fractions were dialyzed at 4 °C into buffer S, concentrated and stored at -80 ºC.

Activation of p38 MAPKα by GST-MKK6b (S207E T211E) – p38 MAPK*α* (10.7 mL, 14 mg) and GST-MKK6b (0.26 mL, 1.6 mg) were incubated for 5 min at 27 °C prior to the addition of ATP (4 mM) in 50 mL of activation buffer B and re-purified, essentially according to the method of Szafranska et al.[19, 20] The activated, tagless p38 MAPK*α* was stored in buffer S at -80 ºC until further use.

7.3.4 Mass Spectrometry, Photodissociation, and High pH Liquid Chromatography

Mass spectrometric measurements were carried out on a Thermo Fisher Scientific LTQ XL (San Jose, CA) using a laser setup similar to that previously described.[21, 22] Photodissociation was performed using a Coherent ExciStar XS excimer laser (Santa Clara, CA) at 193 nm. Direct infusion experiments utilized an online nanoESI setup using a conductive mini microfilter assembly from IDEX Health and Science (Oak Harbor, WA) coupled to a New Objectives uncoated PicoTip® nanoESI emmiter (Woburn, MA), which was operated at an ESI voltage of 2 kV. Peptides diluted to 10 μM in 50/50 acetonitrile/10 mM ammonium acetate (pH 8) were directly infused by syringe pumping at 300 nL/min.

Liquid chromatography using high pH eluents was carried out using a Dionex UltiMate 3000 system (Sunnyvale, CA) with an Agilent ZORBAX 300Extend-C_{18} column (Santa Clara, CA) (150 × 0.3 mm, 3.5 μm particle size). Eluent A was composed of 10 mM piperidine (~pH 11.5) in water and eluent B 10 mM piperidine (~pH 11.5) in acetonitrile. A 65 min linear gradient from 5% eluent B to 40% eluent B was used at a

flow rate of 4 μL/min. The mitogen-activated kinase mixture was injected at approximately five picomoles of tryptic digest per protein. Data-dependent LC-MS/UVPD was performed as follows: the first event was the full mass scan (*m/z* range of 400 – 2000) followed by ten UVPD events on the ten most abundant ions from the first event. One UV pulse was used per MS/MS scan with a *q*-value set to 0.1. Although the actual irradiation time was only 5 ns (the duration of a single laser pulse) for UVPD, the commercial LTQ software limited the activation period to a minimum of 30 μs. The maximum ion injection time was 100 ms for both full scan and MS/MS events, and the ESI voltage was -4 kV. A dynamic exclusion duration of 50 s was used with a list size of 500 allowed m/z values, and a single repeat count. For LC-MS/UVPD, each full mass scan and tandem mass spectrum was the average of three and ten microscans, respectively. Peaks were picked using a signal-to-noise threshold of three and an observable isotopic distribution to assign charge states. For a-EPD experiments, a *q*-value of 0.1 and an activation of 9 pulses (18 ms), 1 mJ/pulse, 500 Hz were used to generate the charge reduced, radical species, which was subsequently activated using a normalized collision energy of 35%, *q*-value of 0.25, and collision duration of 30 ms.

7.4 Results and Discussion

7.4.1 193 nm UVPD of Singly and Multiply Charged Anions

CID of peptide anions generates spectra that are dominated by neutral losses from both precursor and product ions,[8] yielding unforeseen fragmentation behavior and insufficient peptide sequence information. These downfalls are illustrated in **Figure 7.1** for CID of the doubly deprotonated aquaporin peptide, RQpSVELHSPQSLPR. Coupling an ArF excimer laser to a linear ion trap mass spectrometer allows implementation of ultraviolet photodissociation (UVPD) at 193 nm as an alternative to CID (details on the experimental setup can be viewed in the supporting information), yielding rich and informative spectra with high sequence coverage from predictable fragmentation patterns (see **Figure 7.2A**). As seen in this UVPD spectrum of the doubly charged aquaporin peptide, the main sequence ions observed are from the complementary *a*/*x* series, which are present in high abundance and large numbers. Occasionally, hydrogen migration occurs in conjunction with the formation of these *a/x* products, leading to hydrogen-rich or hydrogen-deficient ions denoted as $a_n/x_n \pm 1$ in the spectral figures. The conventional *a*/*x* ions are more consistently observed. A few *c* ions and side-chain loss ions (e.g., *d*, *v*, ions) are also seen in the spectrum as well as *Y*-type ions, which originate from backbone cleavages N-terminal to proline; *Y* ions are produced from radical *x*-ions and have been observed previously upon vacuum UVPD of peptide cations.[23] The combination of these sequence ions originating from both N- and C-termini resulted in 100% sequence coverage, and also allowed confident assignment of the site of phosphorylation. An interesting characteristic of the spectrum shown in **Figure 7.2A** is the highly abundant charge- reduced, radical species (labeled as $[M - H]^{-\bullet}$), which results from photo-induced

electron detachment from the selected peptide precursor. The observation of these types of charged-reduced products was reported previously upon UVPD at 262 nm; subsequent application of CID to these charge-reduced products (termed a-EPD) was used to gain additional peptide sequence information.[17] An example a-EPD spectrum (using 193 nm photoexcitation) is shown in **Figure 7.3**.

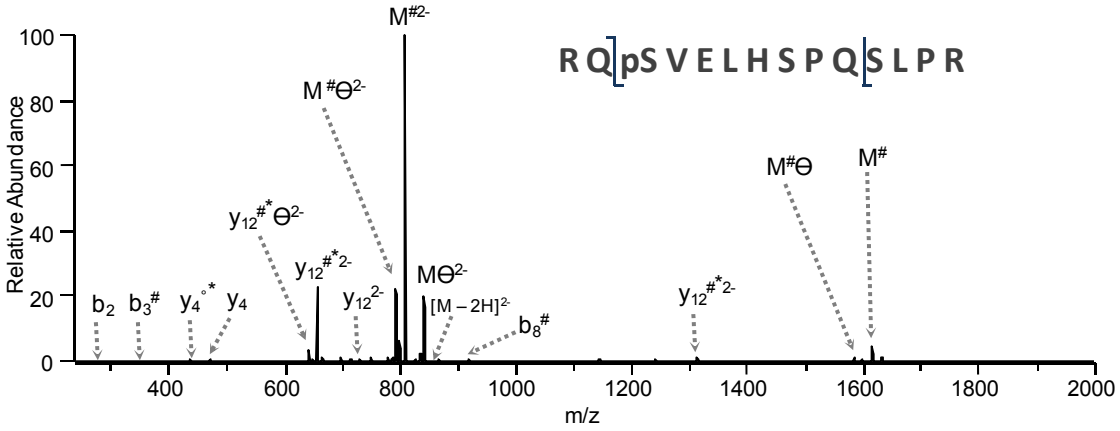

Figure 7.1 CID spectra of human aquaporin-2 peptide, RQpSVELHSPQSLPR (2-). A collision energy of 69 mV, *q*-value of 0.25, and collision duration of 30 ms was used for activation. Neutral losses of phosphoric acid, H_2O, NH_3, and H_2CO are denoted by $^{\#}$, $^{\circ}$, *, and Θ, respectively.

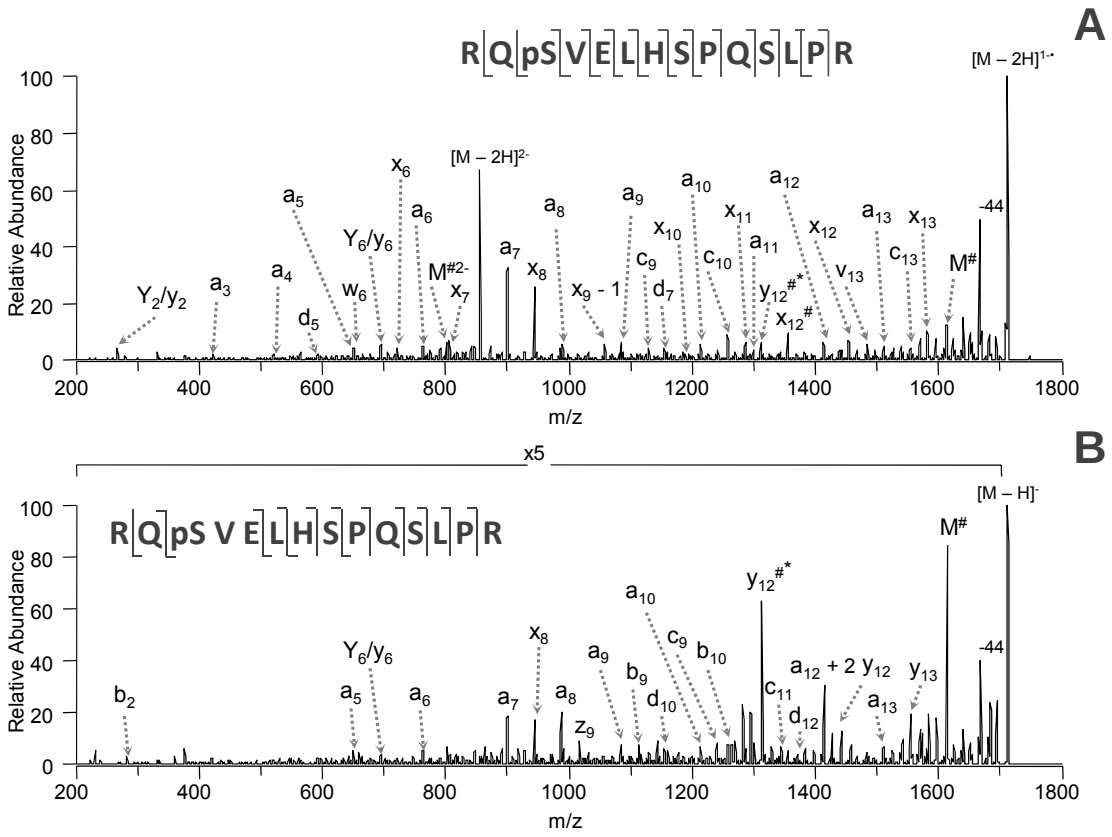

Figure 7.2 UVPD (193 nm) spectra of human aquaporin-2 peptide (**a**) RQpSVELHSPQSLPR, 2- and (**b**) RQpSVELHSPQSLPR, 1-. A *q*-value of 0.1 and an activation of 9 pulses (in a 18 ms activation period), 1 mJ/pulse, at 500 Hz were used for each spectrum. Neutral losses of phosphoric acid, H_2O, and NH_3 are denoted by $^{\#}$, $^{\circ}$, and *, respectively.

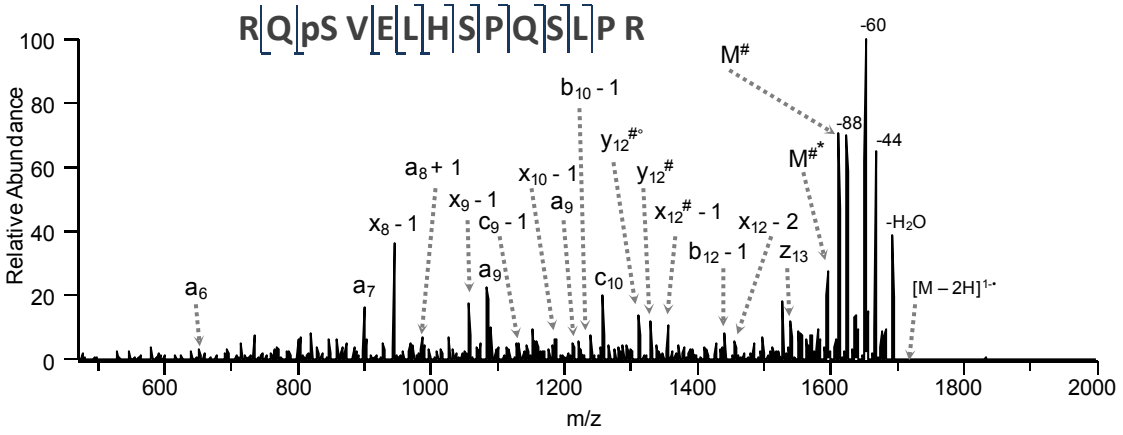

Figure 7.3 Activated – electron photodetachment dissociation (a-EPD) spectra of human aquaporin-2 peptide, RQpSVELHSPQSLPR, 2-. A *q*-value of 0.1 and an activation of 9 pulses (in a 18 ms activation period), 1 mJ/pulse, 500 Hz were used to generate the charge reduced, radical species, which was subsequently activated using a collision energy of 131 mV, *q*-value of 0.25, and collision duration of 30 ms. Neutral losses of phosphoric acid, H_2O, and NH_3 are denoted by $^{\#}$, $^{\circ}$, and *, respectively.

Because ion activation/dissociation is often charge-state dependent, comparisons of the UVPD spectra were undertaken for peptide anions in different charge states**.** **Figure 7.2B** shows the UVPD spectrum of singly charged RQpSVELHSPQSLPR compared to the UVPD spectra of the doubly charged ion in **Figure 7.2A**. The sequence coverage is slightly lower (85%, corresponding to 11/13 possible backbone cleavages) for the 1- charge state than that obtained for the 2- charge state (100% sequence coverage). Interestingly, a very abundant product ion, $y_{12}^{\#*}$, is observed in **Figure 7.2B** and is attributed to the combined loss of phosphoric acid and ammonia from the y_{12} sequence ion. This product arises from the backbone cleavage adjacent (N-terminal side) to the phosphorylated serine residue, suggesting that the highly acidic phosphate group may promote an unusual specific charge dependent cleavage at phosphorylated residues. The same *y*-type ion with phosphoric acid and ammonia loss was observed for UVPD of the

phosphorylated VPQLEIVPNpSAEER peptide as well (data not shown). In this case, an abundant $y_5^{\#*}$ was formed. This preferential pathway could be very useful in pinpointing phosphorylation sites in peptide anions.

Each of the three charge states of the pp60c-src peptide, TSTEPQpYQPGENL, were also subjected to UVPD (**Figure 7.4**). Each charge state yielded many abundant *a*- and *x*-type ions and high sequence coverage (92%, corresponding to 11/12 possible backbone cleavages) as summarized at the top of **Figure 7.4**. The site of phosphorylation could be readily pinpointed due to this high sequence coverage. A particularly salient feature of the UVPD spectrum of the 3- charge state is the decreased low mass cutoff (LMCO), which enables the detection of the abundant PO_3^- and $H_2PO_4^-$ ions. These phosphate reporter anions are important since they permit easy differentiation between phosphorylated and unphosphorylated peptides in complex mixtures,[24] and have recently been used for phosphopeptide mapping of melanoma cells.[25] The singly charged peptide again shows preferential cleavage adjacent (N-terminal side) to the phosphorylated residue as noted by the presence of the $y_7^{\#}$ ion. However, this cleavage is significantly less pronounced and is not accompanied by the neutral loss of ammonia as observed for the RQpSVELHSPQSLPR and VPQLEIVPNpSAEER peptides described above, an occurrence that may be attributed to the differences in fragmentation behavior between phosphorylated serine and tyrosine residues upon UVPD. From these TSTEPQpYQPGENL spectra, it is concluded that significant sequence information and predictable dissociation patterns can be achieved using 193 nm photodissociation for a variety of peptide anion charge states, making it a viable option for ESI-LC-MS/MS applications.

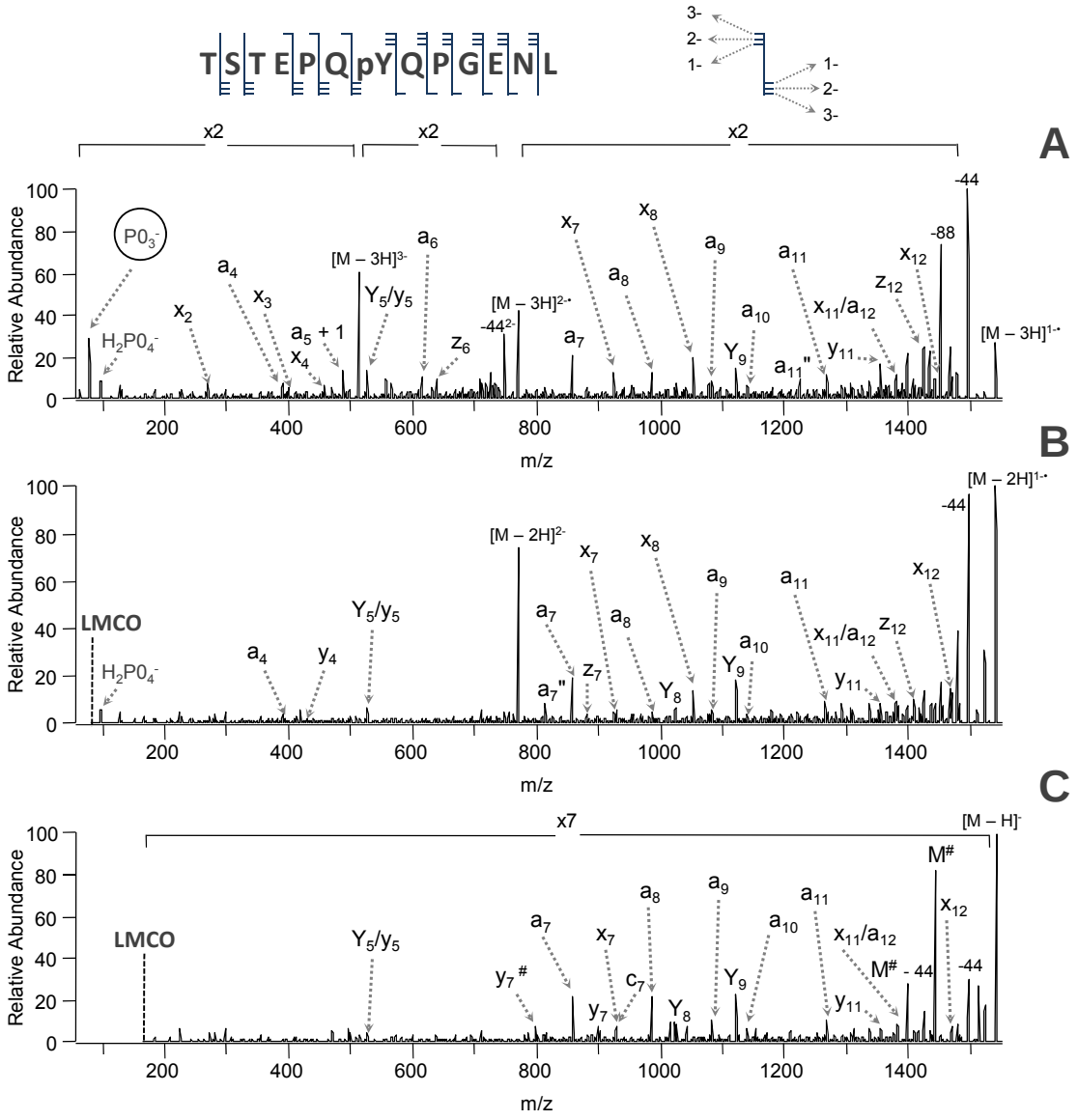

Figure 7.4 UVPD (193 nm) spectra of pp60c-src peptide (**a**) TSTEPQpYQPGENL, 3-, (**b**) TSTEPQpYQPGENL, 2-, and (**c**) TSTEPQpYQPGENL, 1-. A *q*-value of 0.1 and an activation of 9 pulses (in a 18 ms activation period), 1 mJ/pulse, 500 Hz were used for each spectrum. Neutral losses of phosphoric acid and CO_2 are denoted by $^{\#}$ and ", respectively. The legend in the upper right corner represents sequence ion coverage from the 3-, 2-, and 1- precursor charge states.

7.4.2 193 nm UVPD Sequence Coverage Distribution

The production of high abundance sequence ions across the backbone of a peptide is vital for any MS/MS strategy that will be applied to the analysis of complex mixtures. Sequence coverage distribution is defined as the distribution of backbone cleavages based on the relative abundances of complementary *a*/*x* ions, and it was compared for five peptides in the 2- charge state (**Figure 7.5**). Moreover, we also wanted to assess the sequence coverage distribution obtained using a single UV pulse (8 mJ) versus multiple UV pulses (9 pulses, 1 mJ/pulse) for photodissociation, the latter which would be anticipated to cause more extensive dissociation of both precursor and product ions. The sequence coverage distribution is shown in **Figure 7.5** for multiple pulse UVPD (**Figure 7.5A**) and single pulse UVPD (**Figure 7.5B**). For each peptide (RQpSVELHSPQSLPR, HQGLPQEVLNENLLR, TSTEPQpYQPGENL, DRVYIHPFHLVIH, and FFVAPFPEVFGK), the sequence coverage distribution was normalized to the most abundance *a*/*x* pair obtained by any of the two activation methods. Because peptides containing aromatic groups have higher photoabsorptivities at 193 nm and thus typically yield the highest photodissociation efficiencies,[22] the peptides were arranged from the least aromatic (top of figure) to the most aromatic (bottom of figure). Comparison of the distributions in **Figure 7.5A** and **Figure 7.5B** illustrate that activation using a singly high energy UV pulse generates similar or often better sequence coverage distribution compared to that obtained upon multiple lower energy laser pulses. An example spectrum using a single 5 ns (8 mJ) pulse for UVPD activation can be seen in **Figure 7.6** for the peptide FFVAPFPEVFGK (2-). The relative abundances of the surviving precursor ions were two to thirty times lower after nine 1 mJ UV pulses than after a single 8 mJ UV pulse. That is, while the most extensive dissociation of the precursor was always

achieved using multiple UV pulses, the highest sequence coverage distribution was in fact more often obtained using a single UV pulse, especially for the peptides that contained the greatest number of tyrosine and phenylalanine residues. Interestingly, the sequence distributions in both **Figure 7.5A** and **Figure 7.5B** also show an enhanced degree of backbone cleavage next to residues that have UV-absorbing aromatic groups in addition to preferential cleavage at histidine and proline residues. The distributions in **Figure 7.5** indicate that the highest sequence coverage distribution is achieved using a single UV pulse, a natural fit for high throughput LC-MS/MS applications.

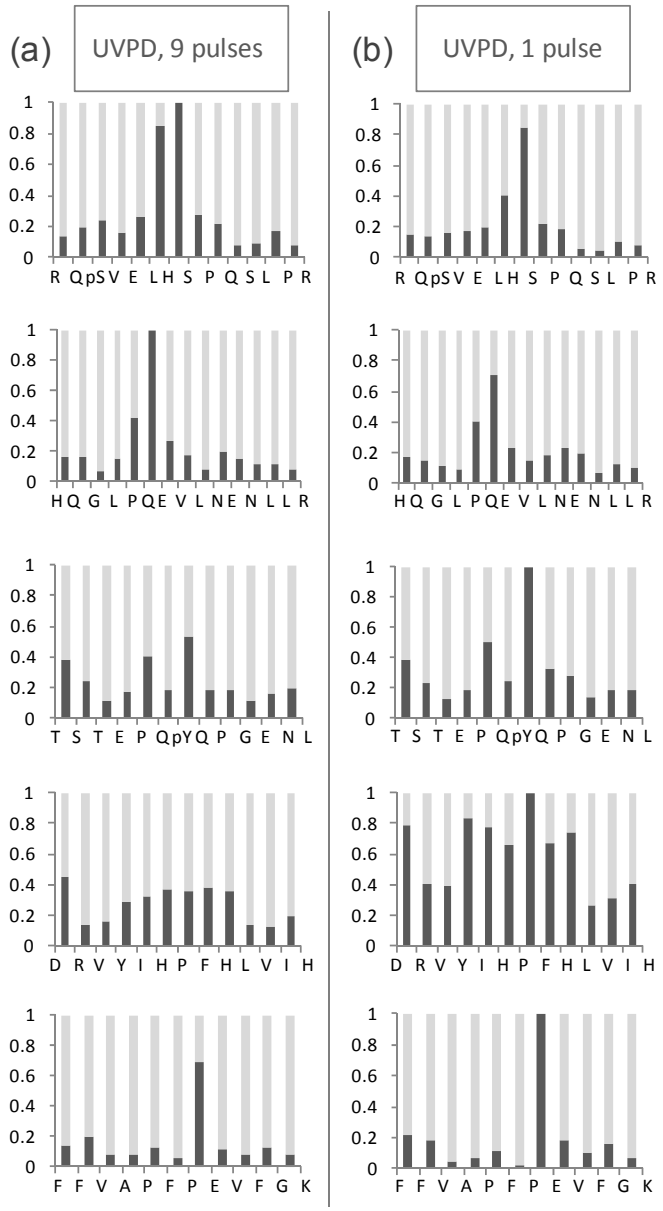

Figure 7.5 Sequence coverage distribution comparison between UVPD using 9 pulses (in a 18 ms period, 1 mJ/pulse) and UVPD using one 5 ns, 8 mJ pulse. The product ion signals were plotted for summed *a*/*x* pairs for each backbone cleavage. For each peptide (RQpSVELHSPQSLPR, HQGLPQEVLNENLLR, TSTEPQpYQPGENL, DRVYIHPFHLVIH, and FFVAPFPEVFGK), ion signals were normalized to the most abundant *a*/*x* pair between the three activation methods. Peptides are oriented from least aromatic (top of figure) to most aromatic (bottom of figure).

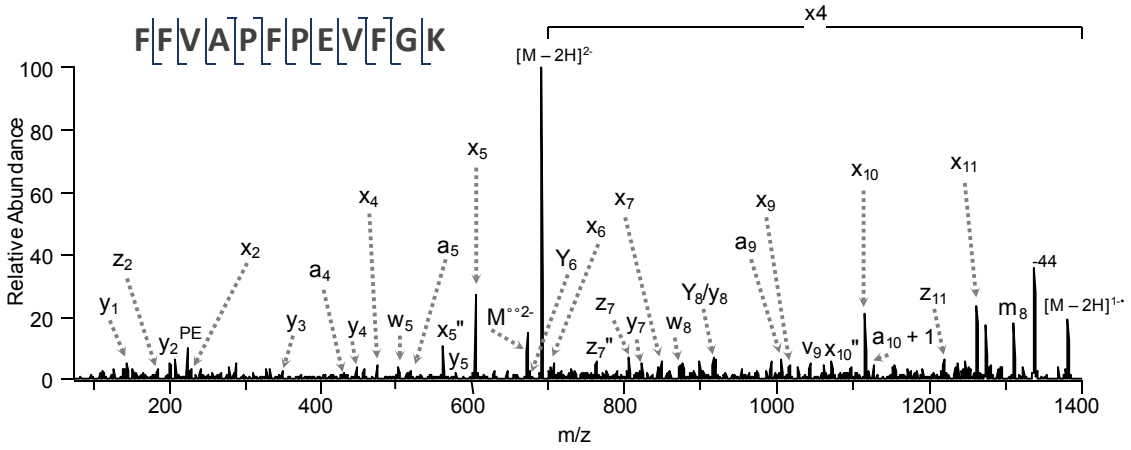

Figure 7.6 UVPD (193 nm) spectra of the α-SI-casein peptide FFVAPFPEVFGK, 2-. A *q*-value of 0.1 and an activation of one 5 ns (8 mJ) pulse were used. Neutral losses of H_2O and CO_2 are denoted by ˚ and ", respectively.

7.4.3 LC-MS/UVPD of Mitogen-Activated Protein Kinases

The promising results summarized above motivated us to adapt UVPD for LC-MS/MS applications, as demonstrated for a mixture of mitogen-activated protein kinases (MAPKs). Typical LC mobile phases used for peptide separations are acidic and limit the ionization efficiencies of acidic species analyzed under negative polarity ionization. Flora and Muddiman showed increased ionization of acidic peptides by adjusting directly infused, working solutions to a higher pH using piperidine.[26] We have replicated these experiments and confirmed increased ion signals as well as a shift to higher charge states upon addition of piperidine to the solutions (data not shown). We were motivated to use piperidine as an organic modifier in the LC mobile phase to promote improved negative ionization after peptide separation to complement UVPD. An equimolar mixture of digested MAPKs, including c-Jun *N*-terminal kinase 2 (JNK2), mitogen-activated protein kinase kinase 7 (MKK7), mitogen-activated protein kinase 1 (ERK2), and p38 mitogen-activated kinase 14 (p38MAPKα), was analyzed. The SDS PAGE gel of each MAPK is

shown in **Figure 7.7**. Mitogen-activated protein kinase pathways are well-defined cell signaling modules that regulate many diverse cellular events. These proteins essentially control all cellular processes and thus are thought to play key roles in cancer progression;[27] they also yield many acidic peptides after trypsin digestion making them amenable to the proposed LC-MS/UVPD method. **Figure 7.8A** shows the base peak chromatogram obtained for the MAPK tryptic peptide mixture using a highly basic mobile phase (pH 11.5) consisting of 10 mM piperidine in each eluent. Good separation and satisfactory peak shapes are obtained using this protocol. The analogous data-dependent UVPD acquisitions yielded high quality photodissociation spectra (i.e., high signal-to-noise) by averaging only 10 microscans using a single 5 ns UV pulse per scan as seen in **Figure 7.8B** and **Figure 7.8C** for the doubly charged p38MAPKα peptide, TLFPGTDHIDQLK, and the triply charged ERK2 peptide, LKELIFEETAR, respectively. Both spectra display mainly *a*/*x* sequence ions along with a few abundant *d* and *w* side-chain loss ions and other diagnostic ions (occasional *Y*, *c, z, y,* and *b* ions); high backbone sequence coverage (100% for both peptides) was derived from this rich fragmentation behavior. These examples illustrate the high sensitivity and sequence information achievable by using UVPD (5 ns activation) on a fast chromatographic timescale.

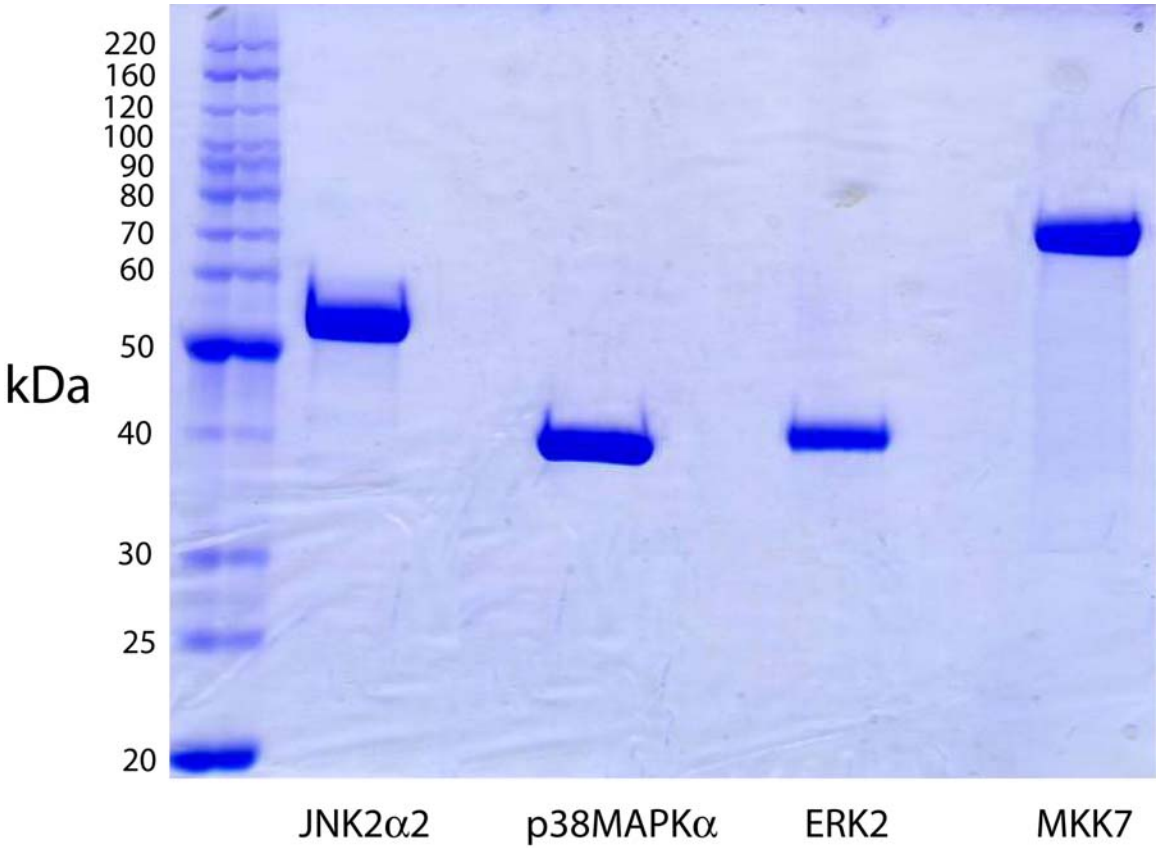

Figure 7.7 SDS PAGE gel of MAPK samples. 10% SDS-Polyacrylamide Gel in tris-glycine running buffer using BenchMark™ Protein Ladder (Invitrogen), Catalog number: 10747-012 showing the degree of purity of active JNK2, p38 MAPKα, ERK2 and GST-MKK7-His_6. The purity of each protein is estimated to be >95%. All purity levels were estimated visually by loading 10-20 μg of each protein.

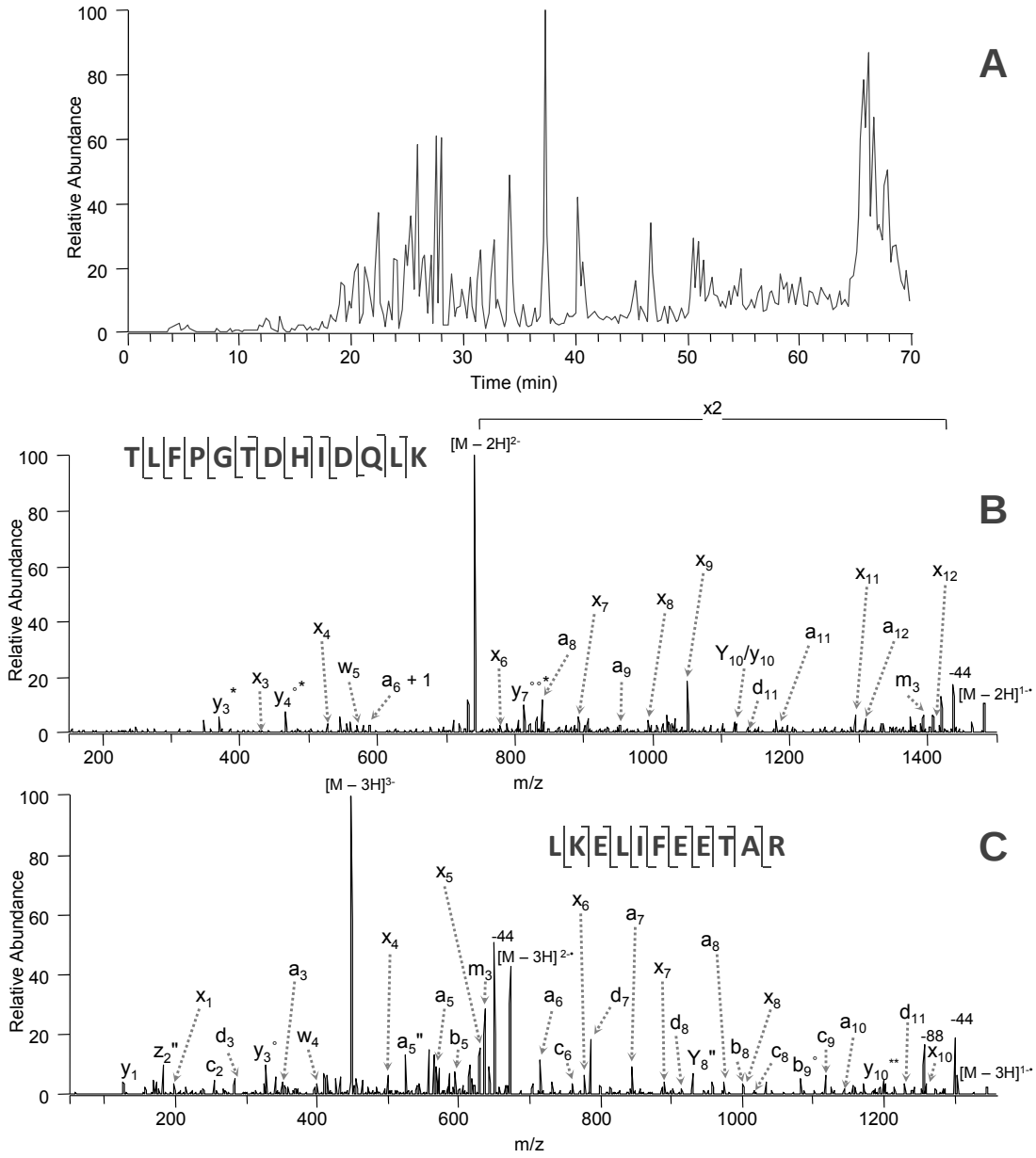

Figure 7.8 LC-MS/UVPD analysis of a mitogen-activated protein kinase (MAPK) mixture; (**a**) base peak ion chromatogram from a separation using 10 mM piperidine spiked into LC eluents, (**b**) UVPD (193 nm) spectra of the p38MAPKα peptide TLFPGTDHIDQLK, 2-, and (**c**) UVPD (193 nm) spectra of the ERK2 peptide LKELIFEETAR, 3-. A *q*-value of 0.1 and an activation of one 5 ns (8 mJ) pulse were used for each 193 nm photodissociation spectrum. Neutral losses of H_2O, NH_3, and CO_2 are denoted by °, *, and ", respectively.

7.5 Conclusions

In this chapter, we have illustrated the benefits and applicability of 193 photodissociation for "bottom-up", acidic proteome characterization. The relative ease and affordability of coupling a laser to a commercial mass spectrometer has accelerated our efforts to implement this MS/MS technique on other mass spectrometers (e.g., orbitraps, hybrid quadrupole time-of-flights, etc.) for various large scale and targeted proteomic applications. We are also currently writing script to manipulate the UVPD spectra to make them more applicable to commercial database searching algorithms, which were mainly developed for positive mode analysis using CID. Coupled with our previous work [20], the present findings establish a compelling benchmark for obtaining high sequence information of both acidic and basic peptides via alternating positive/negative UVPD scans in a single chromatographic separation.

7.6 References

(1) Griffiths, W. J.; Wang, Y. *Chemical Society Reviews* **2009**, *38*, 1882-1896.
(2) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annual Review of Biochemistry* **2001**, *70*, 437-473.
(3) Mitchell, P. *Nature Biotechnology* **2010**, *28*, 665-670.
(4) Duncan, M. W.; Aebersold, R.; Caprioli, R. M. *Nature Biotechnology* **2010**, *28*, 659-664.
(5) de Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Froehlich, F.; Walther, T. C.; Mann, M. *Nature* **2008**, *455*, 1251-1254.
(6) Dalpathado, D. S.; Desaire, H. *Analyst (Cambridge, United Kingdom)* **2008**, *133*, 731-738.
(7) Lee, K. A.; Means, G. D.; Patterson, S. D. *Current Proteomics* **2006**, *3*, 249-257.
(8) Ewing, N. P.; Cassady, C. J. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 105-116.
(9) Zubarev, R. A. *Curr. Opin. Biotechnol.* **2003**, *15*, 12-16.
(10) Zubarev, R. A.; Zubarev, A. R.; Savitski, M. M. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, 753-761.
(11) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 9528-9533.
(12) Coon, J. J. *Analytical Chemistry* **2009**, *81*, 3208-3215.
(13) Kjeldsen, F.; Silivra, O. A.; Ivonin, I. A.; Haselmann, K. F.; Gorshkov, M.; Zubarev, R. A. *Chemistry--A European Journal* **2005**, *11*, 1803-1812.
(14) Kjeldsen, F.; Horning, O. B.; Jensen, S. S.; Giessing, A. M. B.; Jensen, O. N. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, 1156-1162.
(15) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 880-882.
(16) Huzarska, M.; Ugalde, I.; Kaplan, D. A.; Hartmer, R.; Easterling, M. L.; Polfer, N. C. *Analytical Chemistry* **2010**, *82*, 2873-2878.
(17) Antoine, R.; Joly, L.; Tabarin, T.; Broyer, M.; Dugourd, P.; Lemoine, J. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 265-268.
(18) Larraillet, V.; Vorobyev, A.; Brunet, C.; Lemoine, J.; Tsybin, Y. O.; Antoine, R.; Dugourd, P. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 670-680.
(19) Szafranska, A. E.; Dalby, K. N.*FEBS J* **2005**, *272*, 4631-4645.
(20) Szafranska, A. E.; Luo, X.; Dalby, K. N *Analytical Biochemistry* **2005**, *336*, 1-10.
(21) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80*, 4807-4819.
(22) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *Journal of Proteome Research* **2010**, *9*, 4205-4214.

(23) Zhang, L.; Reilly, J. P. *Analytical Chemistry* **2009**, *81*, 7829-7838.
(24) Carr, S. A.; Huddleston, M. J.; Annan, R. S. *Analytical Biochemistry* **1996**, *239*, 180-192.
(25) Old, W. M.; Shabb, J. B.; Houel, S.; Wang, H.; Couts, K. L.; Yen, C.-y.; Litman, E. S.; Croy, C. H.; Meyer-Arendt, K.; Miranda, J. G.; Brown, R. A.; Witze, E. S.; Schweppe, R. E.; Resing, K. A.; Ahn, N. G. *Molecular Cell* **2009**, *34*, 115-131.
(26) Flora, J. W.; Muddiman, D. C. *Analytical Chemistry* **2001**, *73*, 3305-3311.
(27) Sebolt-Leopold, J. S.; Herrera, R. *Nature Reviews Cancer* **2004**, *4*, 937-947.

Chapter 8

Conclusions

Advancing the understanding of biological processes by developing novel tandem mass spectrometry techniques is the key to this dissertation. Tremendous progress has been made in proteomics research due to technological advances in mass spectrometric instrumentation and ion activation; however, there still remain many unresolved limitations in practical applications. One key problem is the inadequate activation of selected precursor ions to generate useful products that are needed to identify proteins and any possible protein modifications. Another important limitation is the high false positive rate in computer-guided interpretation of tandem mass spectra. Both of these problems cause huge limitations in the characterization of the full proteome by MS methods, and the research herein has sought to address these restrictions.

In Chapter 3, precursor ions were supercharged with *m*-nitrobenzyl alcohol (m-NBA) prior to IRMPD analysis. These highly charged species yielded significantly more diagnostic sequence ions when compared to the dissociation of lower charge states. CID of supercharged ions did not yield more sequence information, and in general CID yielded less diagnostic information when compared to IRMPD of high charge states.

Chapter 4 showcased the utility of forming an almost exclusive series of *z*-type by combining electron transfer reactions and N-terminal sulfonation. The MS^n simplification technique illustrated great potential for facile *de novo* sequencing of both unmodified and modified peptides.

IRMPD was implemented onto a dual pressure linear ion trap, and afforded efficient dissociation of intact protein cations as discussed in Chapter 5. Product ions from IRMPD yielded a high level of backbone cleavage specificity and low charge states thus enhancing "top-down" automated database searching using trapping instruments.

The results in Chapter 6 illustrate the utility of applying ultrafast UVPD for high-throughput proteomic workflows. LC-MS/UVPD combined with database searching via SEQUEST improved *in silico* searching results for complex fibrosarcoma cell samples.

The benefits and applicability of 193 photodissociation for peptide anion sequencing and characterization was illustrated in Chapter 7. Upon dissociation by this method, prolific high-energy products (*a*-/*x*-type ions) were generated, often yielding 100% sequence coverage and leaving labile modifications intact. With the use of a basic mobile phase, LC-MS/UVPD was implemented for anion analysis, and resulted in the successful characterization of mitogen-activated pathway kinases.

In summary, this dissertation focused on developing and advancing new mass spectrometry techniques to better aid proteomic analysis. In particular, emphasis was placed on tandem mass spectrometry and MS^n methods to enhance dissociation efficiencies and sequence ion distributions to further aid automated analysis by both database searching (*in silico*) and *de novo* algorithms. The work herein was undertaken on a variety of ion trapping instruments due to their compatibility with lasers for photodissociation and ion/ion reaction capabilities. However, applying these methods to other mass spectrometers with high performance mass analyzers such as quadrupole – time-of-flight and linear ion trap – orbitraps (staples in proteomic analysis) should yield even greater results due the high mass accuracy measurements that can be made. Being that these hybrid mass spectrometers all have trapping devices on the front end, huge hurdles in implementing the methods developed in this dissertation to these high mass

accuracy instruments are not foreseen. The benefits of high mass accuracy measurements would increase the selectivity of automated spectral interpretation by decreasing false positives and increasing true positives thus escalating the depth of analysis into the proteome.

Future goals involve the development of new algorithms based on database searching and *de novo* sequencing to effectively automate spectral interpretation from data generated from the methods developed in this dissertation, those which often generate novel fragmentation behavior. For example, there are no available computational methods for the automated interpretation of anion MS/MS spectra. Using the framework established in Chapter 7, we have begun developing a new searching algorithm for the interpretation of 193 nm UVPD spectra of peptide anions. To start, we modified a recently introduced searching algorithm, MassMatrix,[1] to allow database searching of negative polarity MS/MS files. The computational package uses a mass accuracy sensitive probability-based scoring scheme for both the total number of matched product ions and the total abundance of matched products, and also utilizes LC retention times to filter false positive peptide matches.[2] We have begun to modify the algorithm to be specific to the fragmentation behavior generated from 193 nm UVPD of peptide anions, but also to have the flexibility to search spectra from other photon/electron-based MS^n methods (e.g., NETD, EDD, a-EPD, etc.). To increase the speed and selectivity of the searches, we have incorporated a charge-state-filtering algorithm, which identifies the charge state of each MS/MS spectrum based off the fragmentation patterns, and encodes it into the given mzXML file, a general mass spectrometry data format, prior to searching. The new version of MassMatrix can not only analyze both positive and negative polarity LC-MS/MS files separately, but can combine files from different polarities and different dissociation methods into a single search, thus maximizing the

information content for a given proteomics experiment. MassMatrix will be available as a free search engine to the public and as open-source software at www.massmatrix.net. Very recently we have begun showcasing the utility of combining MassMatrix searching with +/- polarity MS/MS switching (e.g., data-dependent positive MS/MS and negative MS/MS during a single proteomic LC-MS/MS run). **Figure 8.1** illustrates a +/- polarity MS/MS switching experiment in which positive mode ETD was combined with negative mode UVPD in a single LC-MS/MS run of a tryptic digestion of the cytosolic section of HeLa cells. **Figure 8.1A** and **Figure 8.1B** shows the negative and positive base peak chromatogram, respectively, extracted from the single LC-MS/MS file. **Figure 8.1C** illustrates the 193 nm UVPD spectrum of pyruvate kinase isozyme peptide GDLGIEIPAEK (2-) identified from the complex HeLa sample. **Figure 8.1D** shows the ETD spectrum of the triosephosphate isomerase peptide KQSLGELIGTLNAAK (3+) identified from the same LC-MS/MS run of the same HeLa sample. This example illustrates the potential of combining different polarity MS/MS methods in a high-throughput fashion to increase the diagnostic content of a single proteomic analysis. ETD, for instance, has a bias towards yielding the best results for highly basic peptides, while negative UVPD works well with more acidic peptides. By combining the two, it may be possible to yield very high protein sequence coverage. There is also potential for +/- switching just using 193 nm UVPD alone. The fast activation times of UVPD (as low as 5 ns), which are significantly faster than any other activation technique, is an attractive option for elevating the information that can be obtained in a single experiment.

To take advantage of the negative UVPD and/or +/- polarity MS/MS switching methods, the enzyme used for protein digestion is critical. Trypsin is considered the gold standard and generates Arg- or Lys-terminated peptides that have proven well-suited for characterization by CID. While collisional dissociation and tryptic peptides work well

together, negative UVPD would likely match better with other enzymes besides trypsin. For example, the use of Glu-C, which cleaves after acidic residues (D,E), and Asp-N, which cleaves only after aspartic acid (D), would generate peptides that contain acidic residues and thus should enhance negative ESI. Also, chymotrypsin results primarily in peptides terminated with an aromatic residue, a feature that also ensures the peptides will have enhanced UV photoabsorptivity.[3] Coupling the use of these alternative proteases with UVPD and comparison to CID and ETD offers a way to enhance sequence coverage across the basic and acidic proteomes. A streamlined assessment of these alternative enzymes could easily be performed using the MassMatrix algorithm described in the previous section.

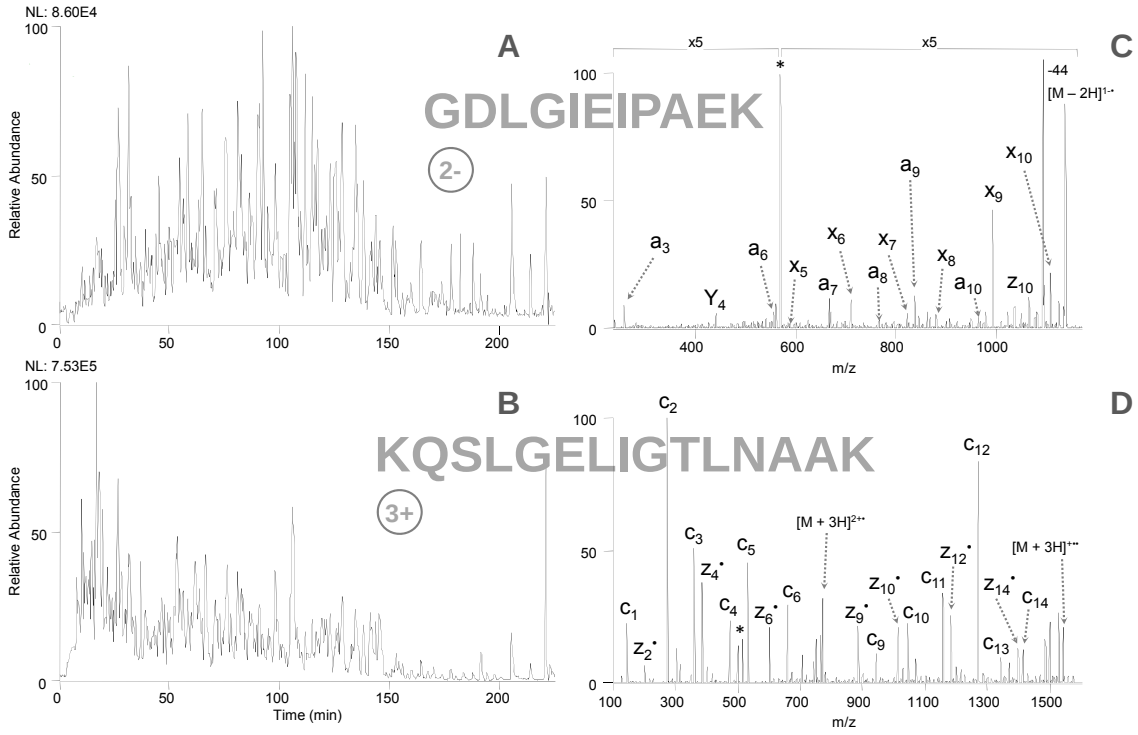

Figure 8.1 Polarity switching LC-MS/MS for simultaneous sequencing of cation (ETD) and anion (193 nm UVPD) tryptic peptides from the cytosolic section of HeLa cells. (a) Extracted negative polarity base peak chromatogram. (b) Extracted positive polarity base peak chromatogram. **c)** 193 nm UVPD of pyruvate kinase isozyme peptide GDLGIEIPAEK, 2-. d) ETD of triosephosphate isomerase peptide KQSLGELIGTLNAAK, 3+.

Another goal for the future also involves expanding the work developed in Chapter 7 of this dissertation, but for intact protein characterization (i.e., top-down analysis). That is, intact proteins would be ionized in the negative mode, subjected to 193 nm UVPD, and then the resulting fragmentation patterns analyzed by search algorithms. Typically, top-down analysis is undertaken in positive MS mode; however, we have recently shown that proteins can be ionized in the negative mode with ion signals on par

with the traditional positive mode analysis. The formation of negative ions is aided by adding approximately 10 – 20 mM of piperidine to working solutions in order to shift the pH to between 11 – 12, thus enhancing deprotonation. A full ESI negative mass spectrum of the protein ubiquitin is shown in **Figure 8.2**, and illustrates the clean and highly abundant signals that can be obtained by this mode of analysis.

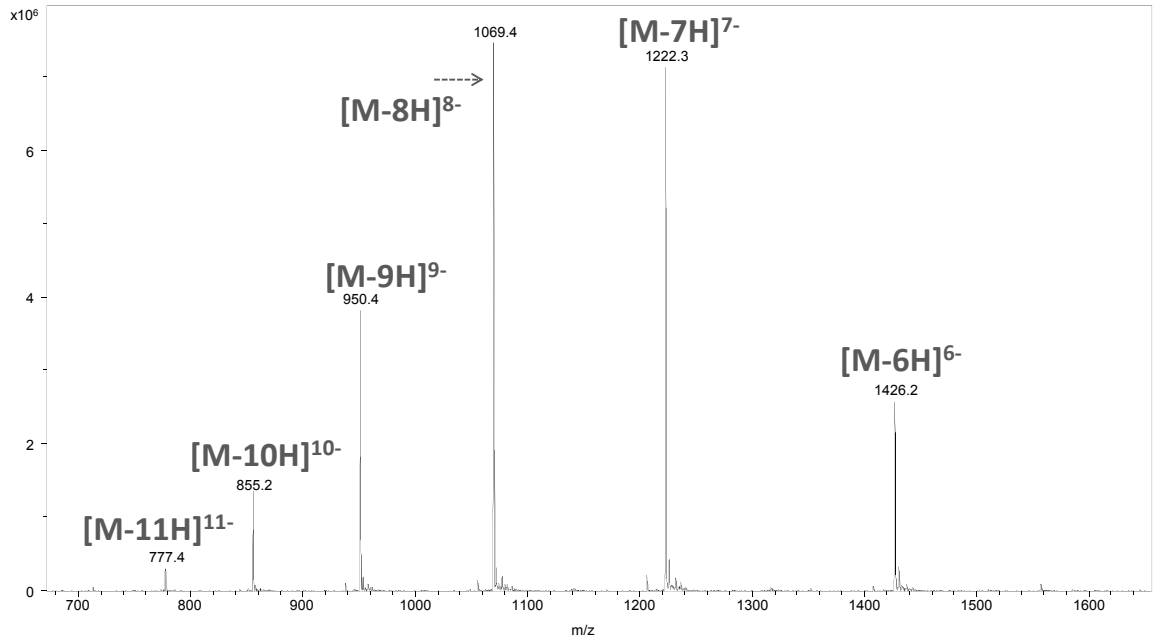

Figure 8.2 Full ESI negative mass spectrum of a ubiquitin solution containing 20 mM of piperidine.

Dissociation of the individual intact protein anion charge states would be performed by 193 nm UVPD to produce diagnostic ions. As mentioned previously, although top-down analysis is more challenging and less developed compared to bottom-up strategies, dissociation of intact proteins offer certain benefits including the elimination of enzymatic digestion procedures, access to contextual post-translational

modification information, and direct knowledge about the molecular weights of the intact proteins.[4, 5] One of the challenges of top-down analysis is generating enough diagnostic information after dissociation. **Figure 8.3A** shows the MS/MS spectrum of intact ubiquitin (7- charge state). As observed in Chapter 7, electron photodetachment, which creates charge-reduced, radical species, is a prominent pathway during UVPD. The sequence ions are generally observed at lower abundance compared to the photodetachment products. **Figure 8.3B** illustrates a zoomed in section from approximately *m/z* 870 – 1180 showing the product ion types matched to peaks in the MS/MS spectrum, and demonstrates the rich information that can be obtained from top-down analysis using 193 nm photons. The sequence information obtained from the negative UVPD dissociation of the intact ubiquitin protein (7- charge state) is shown in **Figure 8.3C**. The colored markers represent cleavage sites along the protein backbone, and are matched to the product ion type by the legend at the right of the figure. As seen in this distribution map, these initial experiments show that high sequence coverage can be obtained by the top-down negative UVPD method. Furthermore, the initial analysis was performed on a fairly low resolution ion trap mass spectrometer limiting the resolving power to approximately five charges on a product ion. If the dissociation technique was implemented on a higher performing instrument such as a q-TOF or orbitrap instrument, a near future goal, a significantly gain in diagnostic information would be predicted. Future goals of this method would be to extend it to higher performing instruments, as described previously, develop automated searching techniques to interpret the fragmentation behavior (likely based off our modified MassMatrix algorithm), and to analyze significantly larger proteins and ones with extensive PTM modifications.

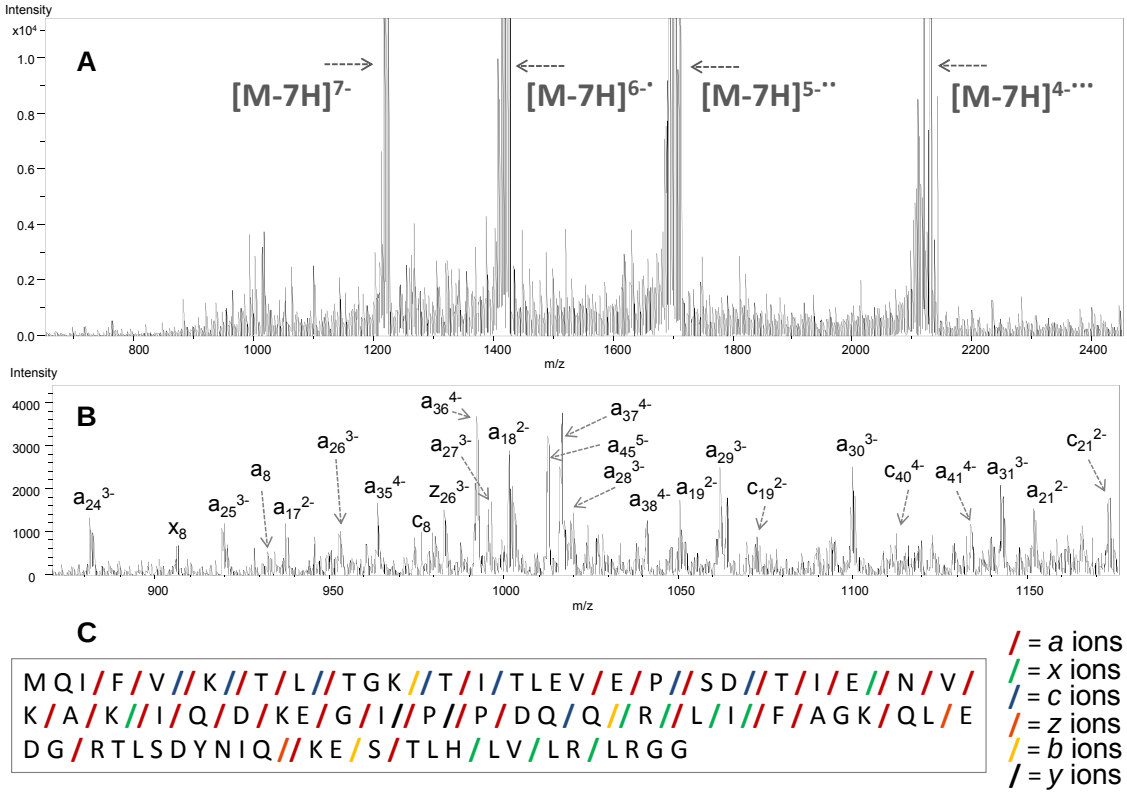

Figure 8.3 Negative polarity top-down analysis of ubiquitin (7- charge state) by 193 nm UVPD. (a) The MS/MS spectrum illustrating the electron photodetachment, radical products (labeled) and the sequence ions at lower abundance. (b) A zoomed in section from approximately *m/z* 870 – 1180 showing the product ion types for peaks in the spectrum. (c) Sequence coverage from the various products observed in the MS/MS spectrum. The legend at the right side of the figure matches the colored backbone cleavage markers with their associated product ion type. The ion types in the legend are listed from most prevalent (top) to least prevalent (bottom).

References

(1) Xu, H.; Freitas Michael, A. *BMC bioinformatics* **2007**, *8*, 133.
(2) Xu, H.; Yang, L.; Freitas Michael, A. *BMC bioinformatics* **2008**, *9*, 347.
(3) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *Journal of Proteome Research* **2010**, *9*, 4205-4214.
(4) Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.
(5) Liu, J.; Huang, T.-Y.; McLuckey, S. A. *Analytical Chemistry* **2009**, *81*, 1433-1441.

Vita

James Andrew Madsen was born in Quincy, Illinois on October 21st, 1982 to Marsha S. Madsen and Robert L. Madsen. After graduating from Quincy High School in the spring of 2001, he attended the University of Colorado at Boulder in the fall of 2001. He graduated *magna cum laude* with a bachelor's degree in chemistry in May of 2005. During his time at the University of Colorado, he helped develop and characterize inhalable therapeutics including the initial work on an inhalable dry powder measles vaccine for impoverished countries under the guidance of Professor Robert E. Sievers. He then moved to Chicago in May of 2005 and worked for Baxter Healthcare Corporation as an analytical chemist in the division of pharmaceutical development up until the fall of 2007. Currently, he has been conducting research in the fundamentals of gas-phase ion chemistry, and developing mass spectrometric-based methods for proteomic and lipidomic applications under the supervision of Professor Jennifer S. Brodbelt at the University of Texas at Austin.

Permanent Address: 2708 East 22nd Street, Austin, TX 78722

This dissertation was typed by the author.